The KDEL Receptor

Structural & Functional Investigations

Philipp Bräuer

Prof. Simon Newstead

Prof. Francis Barr

DPhil Structural Biology

Green Templeton College

If you try and take a cat apart to see how it works, the first thing you have on your hands is a nonworking cat.

– Douglas Adams

Acknowledgements

There are a great number of people who helped, encouraged and influenced the writing of this thesis. My family and friends kept me sane, and without them and their willingness to endure me, I surely would have failed. My supervisors, colleagues and collaborators all played their parts in shaping my thesis and I thank them for all their hard work, ideas and patience. In particular I would like to thank Simon Newstead for his excellent supervision, training, ideas and the chance to work in his lab. And Francis Barr for his supervision, helpful discussions and assistance with the cell biology, and in the writing of our joint paper.

Some people contributed specific work to the contents of this thesis:

- Simon Newstead: Worked together at the beam line collecting data, fishing crystals, building the initial *Mm* KDELR model after MR and providing the finishing touches in refinement.
- Joanne Parker: General supervision and training in the lab and in particular collecting the data on some of the radio-labelled peptide binding assays and initial screening and purification of the human KDELR1.
- Andreas Gerondopoulos: performing, analysing and training in cell biology experiments.
- Iwan Zimmermann: Supervising and assisting the sybody generation experiments in Zurich.
- David Staunton: Performed all the SEC-MALS experiments.

To all those who had their roles to play, large or small, I am grateful.

Abstract

The integrity of the eukaryotic secretory pathway is maintained through selective export and retrieval of proteins between the endoplasmic reticulum (ER) and Golgi. An essential component of ER protein retrieval is the KDEL receptor (KDELR), an integral membrane protein that recognises a C-terminal KDEL signal sequence (a C-terminal KDEL peptide in mammals and HDEL in plants and *Saccharomyces cerevisiae*) in a pH-dependent process. However, the evolutionary origin of the KDELR, the mechanism through which it binds and releases cargo, and how it interacts with cellular trafficking machinery, has remained elusive.

This thesis presents high resolution crystal structures of the *Gallus gallus* KDELR 2, in several states: one apo structure, three peptide-bound structures, and one structure of the KDELR bound to an antagonistic synthetic nanobody (Sybody). These structures are compared and analysed to answer questions about the KDELR's mechanism of action and likely origin of the KDELR.

In vivo cell localisation and *in vitro* binding assays were developed to examine different mutants of the KDELR and test hypotheses generated from structural data. Synthesising the data, a model for the retrieval of ER resident proteins from the Golgi, by the KDELR, was developed. The model involves pH sensitive binding to cargo proteins, controlled by the protonation state of a key histidine residue which stabilises a hydrogen bond within the receptor. The stability of this hydrogen bond in the bound-state locks the receptor onto its cargo until the pH changes. The specificity of the KDELR for HDEL or KDEL motifs was also explored, and appears to depend on differential interactions of the HDEL and KDEL signals with a tryptophan residue in the binding pocket. Comparison with other membrane proteins suggests how the KDELR likely evolved from an ancestral secondary active transporter, demonstrating that transporter-like architectures may have been adapted to transmit signals across cell membranes.

Contents

1 Introduction

Eukaryotic cells are characterised by their membrane bound nucleus and organelles, enabling the segregation of cellular functions into separate compartments with different physiochemical environments. The formation and maintenance of these membrane bound organelles requires the specific transport of proteins and lipids within the cell. To deliver proteins to their target destinations, cells have developed a complex network of protein targeting and trafficking systems. Soluble and membrane proteins destined for the organelles of the secretory and endocytic pathways, or the extracellular space are targeted to the Endoplasmic Reticulum (ER). In fact, over a fifth of mammalian proteins originate in the ER and pass through the secretory pathway, using vesicular transport mechanisms to move between organelles (**Kanapin et al., 2003; Barlowe & Helenius, 2016**). Not only does the secretory pathway transport proteins to their target destination, these same proteins are also post-translationally modified, assessed for folding and assembly quality and, if necessary, selected for degradation. Two major organelles in this transport system are the ER and the Golgi apparatus, which perform a host of essential cellular functions involving protein processing, trafficking and secretion.

1.1 The ER & Golgi

The ER is the largest membrane bound organelle and starting point for secretory and membrane protein biogenesis. Proteins enter the ER lumen either by co-translational or post-translational translocation across the ER membrane and are generally targeted to the ER by a signal peptide (**Phillips & Voeltz, 2016; Rapoport et al., 2017**). In eukaryotes, Sec61 forms the primary channel through which proteins enter the ER. During co-translational translocation, the ribosome nascent-chain complex is targeted to Sec61 by the signal recognition particle (SRP) and its receptor the SRP receptor (SR), which results in the polypeptide chain being transferred directly from the ribosome to the Sec61 channel (Rapoport et al., 2017). Integral membrane proteins are also co-translationally inserted into the membrane from Sec61 and almost all multi-pass membrane proteins are thought to use the Sec translocon. However, the mechanism by which these are inserted into the membrane by Sec61 with defined topology remains poorly understood, and may require other accessory factors such as the multi-subunit ER membrane protein complex (EMC), to help guide insertion of transmembrane (TM) domains (Guna & Hegde, 2018; Chitwood & Hegde, 2019). Post-translational translocation substrates are not recognised by SRP during synthesis and cytoplasmic factors such as calmodulin, Hsp70 family chaperones and small glutamine-rich tetratricopeptide repeat-containing protein alpha (SGTA) prevent them from adopting a stably folded state before translocation (Guna & Hegde, 2018; Johnson et al., 2013). Because ongoing translation is not a driving force for post-translational translocation, another mechanism is required for these proteins to cross the membrane. Binding immunoglobulin protein (BiP) (a luminal Hsp70 family member) acts as a molecular ratchet to facilitate the transport through Sec61 once the polypeptide has inserted into the translocon (Matlack et al., 1999). Tail-anchored membrane proteins are inserted into the membrane by recognition of the GET/TRC complex which recognises their transmembrane domain, once the TM domain is loaded onto Get3 (TRC40 in mammals), it is captured by the membrane protein Get1-Get2 (WRB/CAML in mammals) receptor complex, which inserts the TM domain into the ER membrane (Hegde & Keenan, 2011; Vilardi et al., 2014). The EMC also appears to have a role in inserting tail-anchored proteins not recognised by GET and additionally appears to co-translationally insert N-terminal TM domains, though how the EMC performs its role remains unclear (Chitwood & Hegde, 2019).

Once inside the ER, proteins are met with a host of other proteins responsible for their folding, post-translational modification and quality control of their folding state. If proteins fail to become folded, they are targeted to ER-associated degradation (ERAD), a process in which misfolded proteins are targetted for ER retrotranslocation, ubiquitination and subsequent degradation by the proteosome (Ruggiano et al., 2014). Only once the proteins are properly folded and thus no longer bound by chaperones do the ER quality control mechanisms permit them to be exported onwards to the Golgi by coat protein complex II (COPII) transport vesicles. For example, proteins in the calnexin/calreticulin cycle, which assists with protein folding, make use of an N-glycan which is recognised by the lectins calnexin and calreticulin and is illustrated in figure 1.1. Once the innermost glucose of the oligosaccharide is trimmed by glucosidase II, the protein is released from calnexin/calreticulin. Subsequently, it is recognised by UDP-glucose/glycoprotein glucosyl transferase (UGGT) which will reglycosilate the glycan if the correct conformation has not been achieved, thus returning the protein to repeat the cycle. Persistently unfolded proteins are eventually degraded, before which their N-glycans are substantially trimmed (Araki & Nagata, 2011). Proteins that are successfully folded are transported onwards towards their destination. In the Golgi, proteins are again met with processing proteins which complete the oligosaccharide modifications, as well as other post-translational modifications including phosphorylation and sulphation.

In order for the ER and Golgi to perform their functions, they are host to a large number of resident proteins, which act inside the luminal environment and must therefore be targeted to and retained within these organelles. There is however a large flow from the ER to Golgi. Roughly 1.5–5% of the ER volume flows out per minute and it has been estimated that this would represent 5–10% of the Golgi volume entering every minute (Thor et al., 2009; Cancino et al., 2013). The flow between ER and Golgi is not only due to the large number of proteins synthesised in the ER that need to travel to their destinations. The ER is also a major site of lipid synthesis in the cell, where a majority of phospholipids and

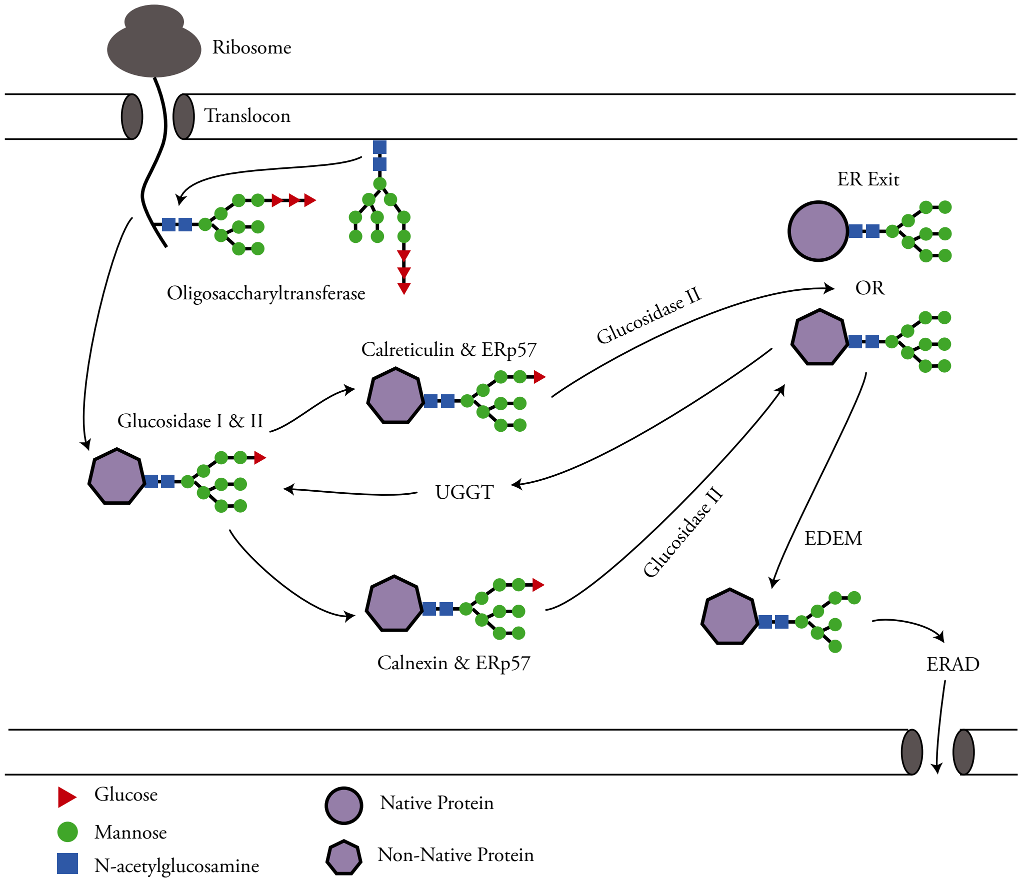


Figure 1.1 The calnexin/calreticulin cycle takes place inside the ER. Newly synthesised glycoproteins are N-glycosylated at an asparagine side chain in the Asn-X-Ser/Thr consensus sequence by oligosaccharyltransferase, which transfers a 14-sugar oligosaccharide from diolichol. The two terminal glucose residues of the oligosaccharide are then trimmed by glucosidase I and II. The resulting oligosaccharide is recognised by calreticulin and calnexin. When bound by these, client aggregation and premature ER export is prevented and folding is assisted by interaction with other folding factors such as ERp57, a thiol-disulfide oxioreductase. Removal of the final glucose by glucosidase II prevents further interaction with calnexin and calreticulin. If the client protein is in a native, folded state it will be permitted to exit the ER. UGGT recognises proteins that are unfolded and can reglycosilate non-native glycoproteins using a UDP-glucose sugar donor, these are then recognised by calnexin and calreticulin and re-enter the cycle. Persistently unfolded proteins will encounter slower acting mannosidases which trim the glycan structure, the resulting α1,6-linked mannose signature is recognised as an ERAD substrate and targeted for retrotranslocation, ubiquitination and degradation by the cytosolic proteosome (Ellgaard & Helenius, 2003; Ellgaard & Frickel, 2003; Clerc et al., 2009; Ruggiano et al., 2014).

cholesterol are produced and rapidly transported to other organelles. Despite being the site for cholesterol synthesis, the ER membrane contains very low concentrations of these, resulting in a more permeable membrane packing that may be beneficial for insertion of proteins and newly-synthesised lipids (van

Meer et al., 2008). Cells are thus thought to employ several active retention and retrieval mechanisms for ER resident proteins that are swept along the ER to Golgi flow, allowing them to maintain stable ER composition, which is critical for normal cell function (Barlowe & Helenius, 2016; Oakes & Papa, 2015).

The traffic between ER and Golgi is mediated by coat protein complex I (COPI) and COPII vesicles. COPII vesicles are involved in protein export from the ER to the Golgi and the ER-Golgi intermediate compartment (ERGIC), a tubovesiuclar membrane cluster thought to be involved with trafficking and cargo sorting between ER and Golgi. COPII components include the inner coat dimer of SEC23-SEC24 and the outer-coat heterotetramer consisting of SEC13-SEC31, shown in figure 1.2 (Barlowe et al., 1994). COPII associates with ER exit sites (ERES) where assembly is initiated by SAR1 activation through the SAR1-Guanine nucleotide Exchange Factor (GEF) (SEC12) (Barlowe et al., 1994). SAR1 is a GTPase and SAR1-GTP acts as a docking platform for SEC23-24 which subsequently is bound by SEC13-31 (Gomez-Navarro & Miller, 2016; Zanetti et al., 2013). Due to the large number of different proteins passing from the ER onwards, there is no single sorting signal. The COPII machinery can recognise a variety of sorting signals including di-acidic ([DE]*x*[DE]), di-hydrophobic and di-aromatic ([FY][FY]) motifs, which generally bind SEC24, an example of which is seen in figure 1.2 (Gomez-Navarro & Miller, 2016; Dell'Angelica & Bonifacino, 2019). In addition, COPII may also recognise conformational or folded epitopes such as that of Sec22, a soluble N-ethylmaleimide-sensitive fusion protein attachment protein (SNAP) receptor (SNARE) protein involved in targeting and fusion of transport vesicles (Mancias & Goldberg, 2007, 2008). Transmembrane COPII cargo proteins can bind the COPII proteins directly, while luminal proteins rely on export receptors with the appropriate cytosolic sorting signals (Barlowe & Helenius, 2016).

COPI vesicles are mainly involved in the retrieval of ER resident proteins as well as cargo receptors from the ERGIC and Golgi to the ER (Dell'Angelica & Bonifacino, 2019). COPI consists of seven subunits, the B-subcomplex of

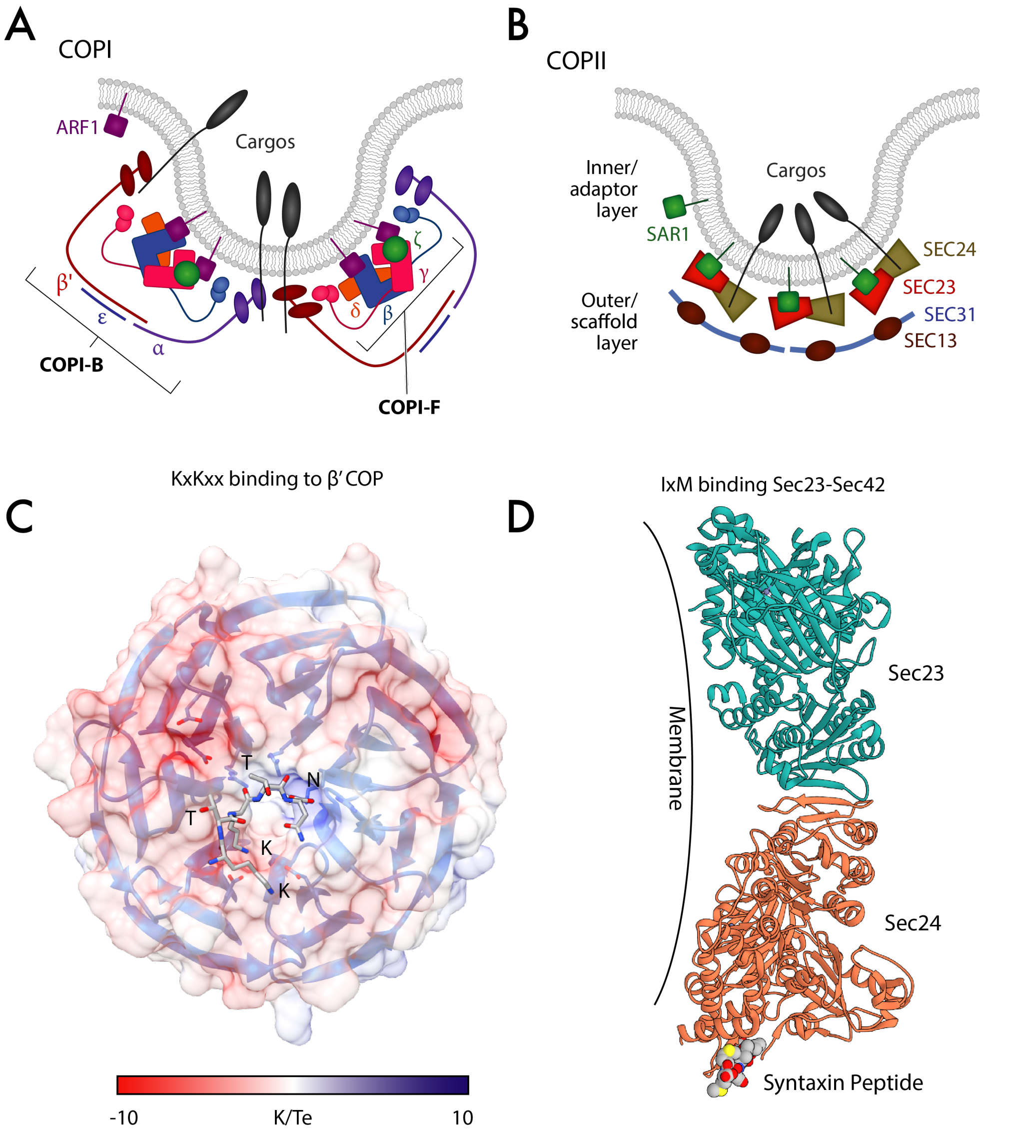


Figure 1.2 **A** shows the cartoon arrangement of a COPI coat with its components, **B** shows the same for COPII. Adapted from Dell'Angelica & Bonifacino (2019). **C** shows the structure of the N-terminal β-propeller of β′-COP (with the electrostatic surface shown (Jurrus et al., 2018)) bound to a K*x*K*xx* motif bearing peptide, which is recognised by several acidic residues on the propeller, the C-terminal carboxy group and -2 and -3 carbonyl groups are also recognised by several positively charged residues (PDB: 2YNP) (Jackson et al., 2012). **D** The structure of COPII Sec23-Sec24 bound to a Syntaxin-5 peptide with an IxM motif. The peptide is bound adjacent to a β-sheet and interacts as another strand, with additional hydrophobic interactions from side-chain interactions. The location of the membrane is indicated by the black line (PDB: 3EFO) (Mancias & Goldberg, 2008).

α-COP, β′-COP and ε-COP, and the F-subcomplex of β-COP, γ-COP, δ-COP and ζ-COP, shown in figure 1.2 (Waters et al., 1991). Assembly of COPI occurs when

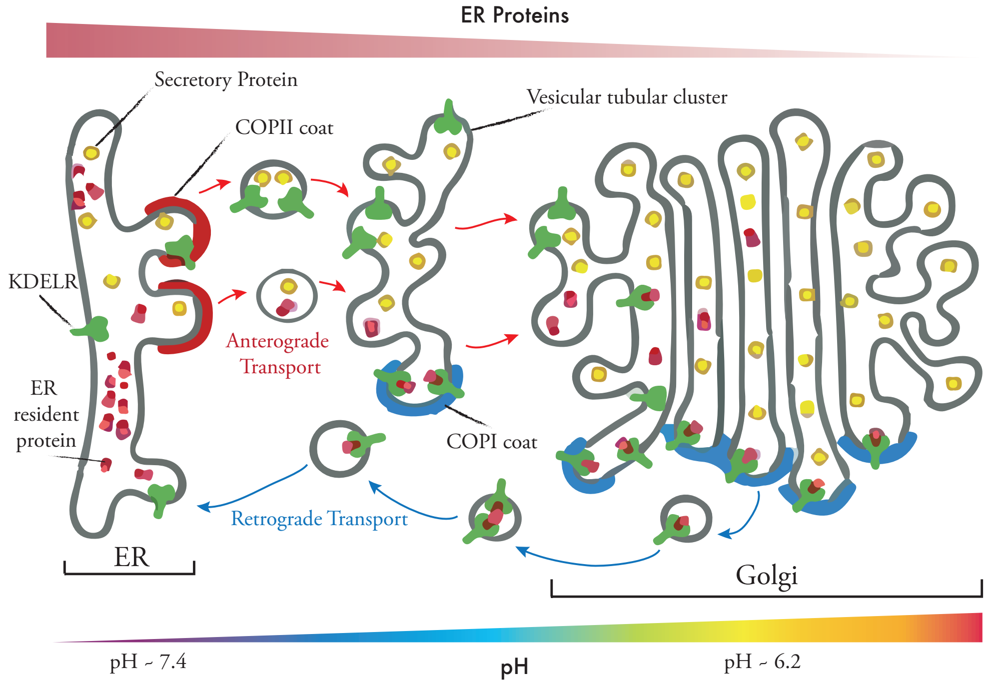


Figure 1.3 The KDELR shown in its cellular context. Resident ER proteins that are found in the Golgi are recognised by the KDELR and returned to the ER via COPI vesicles. The KDELR releases the cargo and returns to the Golgi in COPII vesicles. Concentration of ER proteins and the pH gradient are shown using the coloured triangles. Adapted from Alberts et al. (2007).

GTP bound ARF1, a small membrane associated GTPase which is activated by SEC7 family GEFs, binds COPI by interactions on β-COP and γ-COP.

These COPI-ARF1 complexes form triads that co-assemble laterally to generate a coat at the membrane surface (Dodonova et al., 2015). The best characterised COPI sorting signals are the cytosolic KK*xx* and K*x*K*xx* di-basic motifs, found at the C-terminus of transmembrane proteins (Cosson & Letourneur, 1994; Jackson et al., 2012). These di-basic motifs are recognised by the β-propeller WD40 domain of α-COP (KK*xx*) and β′-COP (K*x*K*xx*), an example of which is shown in figure 1.2 (Ma & Goldberg, 2013). COPI also mediates the recycling of glycosyltransferases to the *cis*-Golgi, from further parts of the Golgi complex via an Ø[KR]*x*L*x*[KR] signal that interacts with δ-COP and ζ-COP (where Ø is L, I, V, M or F) (Liu et al., 2018). Retrieval of other transmembrane proteins to the ER occurs via the help of Rer1, itself a transmembrane protein with an unknown COPI signal (Sato et al., 2003).

Soluble ER resident proteins are retrieved by two pH-dependent mechanisms and returned to the ER by COPI vesicles. The first mechanism involves protein recognition in the Golgi by the KDEL Receptor (KDELR), a membrane protein which recognises ER resident proteins by their eponymous C-terminal peptide motif Lys-Asp-Glu-Leu (K–D–E–L) (**Munro & Pelham**, 1987; **Semenza et al.**, 1990). Figure 1.3 shows a simplified cycle of transport events between the ER and Golgi, and the KDELR's role therein. The second mechanism involves the Erv41-Erv46 complex (both transmembrane proteins), whose recognition signal remains unclear (**Shibuya et al.**, 2015).

The pH gradient in the secretory system, upon which Erv41-Erv46 and KDELR ER retrieval rely, is generated by vacuolar-type H^+ ATPases (V-ATPases). It is thought that the pH gradient is established by a progressively higher concentration/activity of V-ATPases towards the late or *trans* compartments of the secretory pathway, with an increasing proton leak throughout the membranes of the secretory organelles in the opposite direction (**Wu et al.**, 2001; **Paroutis et al.**, 2004). This results in a relatively neutral ER with a decrease in pH as one moves across the Golgi from *cis* to *trans* towards secretory granules or endosomes and lysosomes in the endocytic pathway. When this pH gradient is collapsed in the Golgi, post-translational modifications are hindered and cargo mis-localises (**Paroutis et al.**, 2004).

1.2 The KDEL Receptor

The KDELR is an integral membrane protein, localised to the early or *cis*-Golgi where it can efficiently capture escaped ER luminal proteins (**Griffiths et al.**, 1994; **Semenza et al.**, 1990). The KDELR was first discovered in yeast by **Semenza et al.** (1990), as one of two ER retention deficient (ERD) *Saccharomyces cerevisiae* (*Sc*) mutants. That the KDEL sequence was required and sufficient for ER retention was shown a few years previously by the same group, which demonstrated that removal of the KDEL sequence from BiP in COS cells resulted in the secretion of BiP, while the addition of a KDEL sequence to the C-terminus of lysozyme prevented secretion (Munro & Pelham, 1987). The first evidence that ER proteins are retrieved from a post ER compartment was provided when Pelham (1988) attached a KDEL sequence to cathepsin D, a lysosomal protein whose oligosaccharide side chains are generally processed by post ER enzymes. While cathepsin D tagged with KDEL localises to the ER, its oligosaccharide chains appear to be modified by N-acetylglucosaminyl-1-phosphotransferase, which is found in the *cis*-Golgi, indicating the presence of a retrieval mechanism from post ER compartments. The first human homologue of the KDELR (KDELR1) was also discovered by Pelham's group in 1990 and its localisation to the Golgi and intermediate compartments between ER and Golgi was described, as well as the first prediction of seven transmembrane helixes from hydropathy plots (Lewis & Pelham, 1990). A second human homologue (KDELR2) was published in 1992 (Lewis & Pelham, 1992b). Previously that year, Lewis & Pelham (1992a) had shown that KDEL tagged lysozyme caused a redistribution of the KDELR from the Golgi to the ER while an over expression of KDELR allowed the retention of off-target KDEL sequence tagged proteins.

1.2.1 Physiological Functions of the KDEL Receptor

The ability of the KDELR to carry out its function correctly in a cell has been linked with a variety of cellular processes, diseases and infections. Perhaps most obviously the retrieval of ER resident proteins is important in protein folding stress and ensuring that protein aggregates do not build up. Some ER resident proteins are brought to the Golgi while still attached to their client proteins and mis-folded or unassembled proteins have been shown to cycle between the ER and Golgi, thus involving the KDELR directly in ER quality control (Hsu et al., 1991; Hammond & Helenius, 1994). Upon the disruption of retrograde transport of the KDELR, unassembled TCRα (T-cell receptor), targeted for ERAD under normal conditions, escapes this fate. As these proteins do not contain KDEL motifs it is thought that ER resident proteins are bound to them during this retrieval cycle (Yamamoto et al., 2001). ERp44 is a protein that in coordination with the KDELR, returns immature oligomeric proteins such as IgM and adiponectin, as well as some ER resident proteins that lack a KDELR retrieval motif, from the Golgi to the ER (Watanabe et al., 2017). Like the KDELR, it binds its cargo in a pH-dependent manner which exposes its C-terminal KDEL motif (RDEL) leading to the binding of the KDELR and the retrograde transport of ERp44 with its attached client (Watanabe et al., 2017).

Mammalian genomes encode three KDELRs, in humans these have a pairwise sequence identity between 73.0–83.5% and are seen aligned in figure 1.4. The three human KDELR homologues are widely expressed in different cell types, with KDELR1 generally being transcribed in higher quantities than KDELR2 which in turn is transcribed in higher quantities than KDELR3 (Raykhel et al., 2007). No knockdown data exists for all three KDELRs in mammals, but in *Drosophila*, which only contain one copy (75% identity to human KDELR1), a loss of function mutation is early larval lethal (Abrams et al., 2013). The loss of a KDEL retention motifs in humans, has been associated with myeloproliferative disease (*CALR* gene, encoding calreticulin) and osteogenesis imperfecta (*LEPRE1* gene, encoding Prolyl 3-hydroxylase 1) (Klampfl et al., 2013; Takagi et al., 2012). While mice expressing mutant BiP missing the KDEL motif die within one day of birth due to respiratory failure and display neuronal development defects (Mimura et al., 2007, 2008). Mice expressing the BiP mutant heterozygously develop motor disabilities in ageing (Jin et al., 2014, 2017).

Recessive missense mutation of KDELR1 in mice is associated with lymphopenia (decreased number of lymphocytes in the blood), reduced expression of the TCR, and compromised antiviral immunity (Siggs et al., 2015). Other transgenic mice strains expressing KDELR1 with a point mutation have shown decreased numbers of naïve T-cells, potentially due to the KDELRs role in the integrated stress response (ISR) via protein phosphatase 1 (Kamimura et al., 2015). Interestingly, transgenic mice expressing a mutant human KDELR1 develop dilated cardiomyopathy and aggregates of mis-folded proteins can be found in their cardiomyocytes - and also showed several other pathological and clinical features of human dilated cardiomyopathy (Hamada et al., 2004). In human cell lines KDELR2 and KDELR3 are thought to be up-regulated as part of the unfolded protein response (UPR) in order to prevent loss of ER resident proteins vital in dealing with unfolded proteins (Trychta et al., 2018). ER stress has been implicated in a multitude of human disease pathologies including neurodegeneration, heart disease, and cancer, and can play a role in viral infections, thus directly and indirectly implicating the KDELR in a vast number of diseases (Oakes & Papa, 2015; Lin et al., 2008; Wang et al., 2011; Chen et al., 2019).

Several viruses have been shown to interact with the KDELR. Dengue virus binds KDELR in the ER via its pre-membrane protein and when this interaction is blocked, progeny virus accumulates in the ER and is not properly transported along the secretory pathway (Li et al., 2015). Japanese encephalitis virus has also been shown to make use of the KDELR for correct trafficking of its particles in the secretory pathway, and knockdowns of KDELR result in accumulation of intracellular-assembled virus and a decrease in extracellular viral RNA (Wang et al., 2016). Several African swine fever virus proteins have C-terminal KDEL motifs. These are thought to be associated with the envelopment of virions in a pathway involving ER membranes (Netherton et al., 2004). KDELR binding of ER chaperones is also thought to compete with binding of these to measles virus envelope proteins, decreasing viral titres as chaperones are not available to assist viral protein folding (Tiwarekar et al., 2019).

A/B toxins such as cholera toxin, pseudomonas exotoxin (PE) and killer toxin K28 contain a KDEL C-terminal motif thought to be important in their retrograde transport through the secretory path of target cells. It has been shown that K28 uses the KDELR to enter target cells (Becker et al., 2016). PE contains the C-terminal sequence REDLK which cannot bind the KDELR, removal of the terminal lysine allows binding to occur. Removal of the last residue can occur outside the cell in plasma and appears to be critical in allowing PE to travel to the ER, from where it is thought to translocate to the cytosol (Hessler & Kreitman, 1997; Jackson et al., 1999). In fact, replacing the C-terminal sequence of PE with the canonical KDEL sequence can enhance its cytotoxicity (Seetharam et al., 1991).

The KDELR therefore performs a critical task in eukaryotic cells and its failure or exploitation by viral proteins or toxins leads to catastrophic outcomes for the cell. This highlights the importance of understanding the molecular mechanisms behind the KDELR, which impact such a variety of important cellular processes.

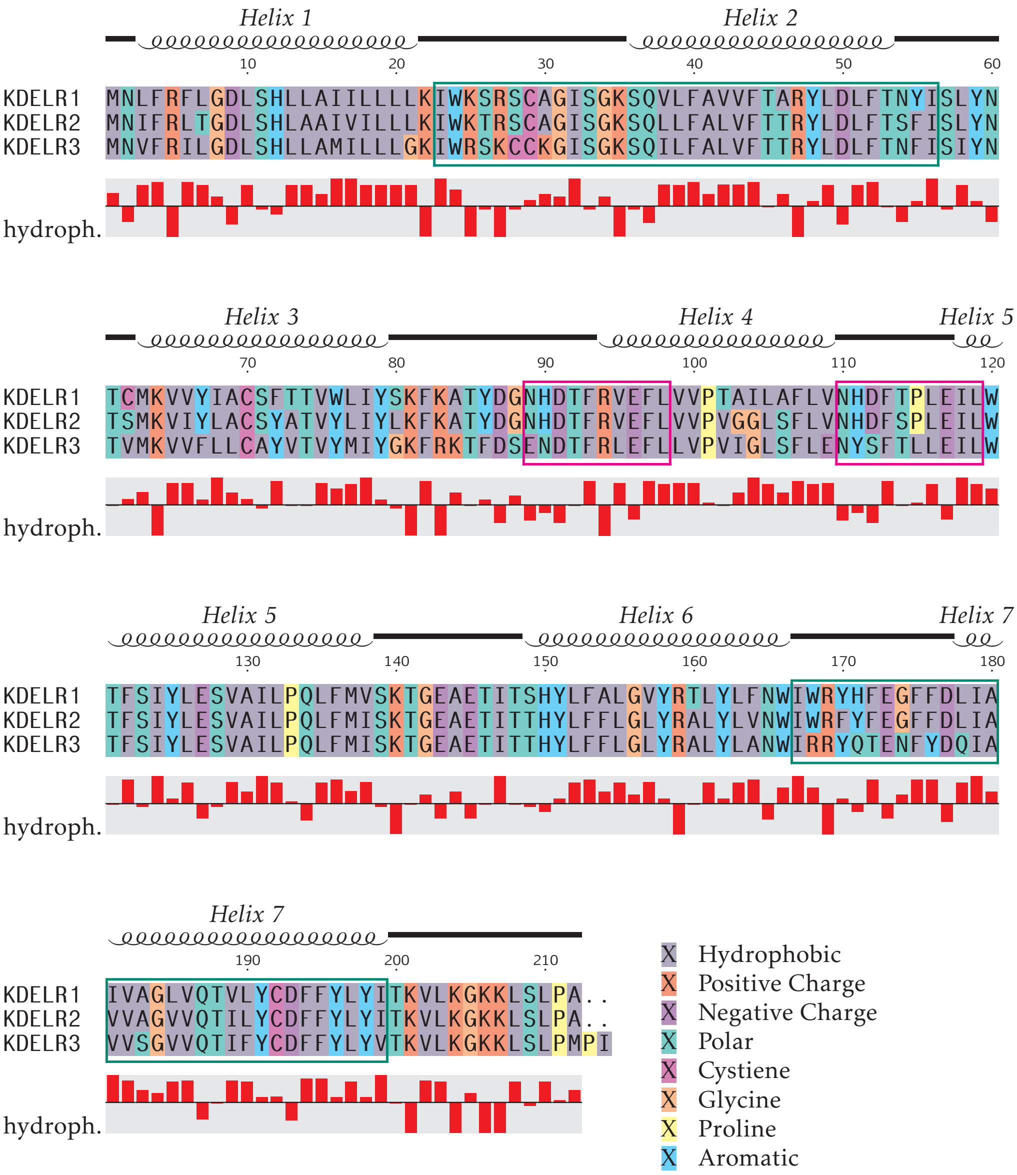


Figure 1.4 Sequence alignment of the three human KDELRs. Shown with predicted secondary structure (average TM predicted by Quick2D (Zimmermann et al., 2018b)), hydrophobicity plot (from (Lesser & Rose, 1990)) and repeats (boxed regions, from RADAR (Heger & Holm, 2000)).

1.2.2 KDEL Receptor and KDEL Cargo Binding

The first *in vitro* evidence of the KDELR binding to a KDEL peptide came when Wilson et al. (1993) described the binding of a radio-labelled KDEL peptide to rat liver Golgi membrane fractions and total membrane extracts of COS cells expressing KDELR1. This also revealed for the first time the pH-dependent binding of the KDELR to its peptide ligand. Scheel & Pelham (1996) reconstituted KDELR into liposomes using various phospholipids and could corroborate the findings that the KDELR binding to the KDEL sequence has a nanomolar dissociation constant (again by incubation with radio-labelled KDEL peptides) and does indeed demonstrate a pH-dependence under these conditions.

This observed pH-dependence led to the proposal that binding of KDEL cargo proteins in the Golgi and their release from the KDELR in the ER was reliant on the pH difference between the two organelles. The pH of the ER has been estimated to be pH 7.2–7.4 while the pH of the *cis*-Golgi is estimated to be around 6.2, with the late Golgi increasing in acidity (Wu et al., 2000, 2001). The pI of a C-terminal KDEL tetrapeptide is calculated to be 5.2, which is close to the reported optimal binding pH of 5 (Scheel & Pelham, 1996). Whether the pH difference between ER and Golgi is sufficient to explain the differential binding and release of KDEL cargo proteins remains an open question. To date, no other energy source for KDEL sorting against the concentration gradient of its cargo has been proposed.

After the discovery of the KDELR and its function as a receptor of the KDEL motif involved in ER retention, focus shifted to discerning its mechanisms of action and structural features. Initial models of the seven transmembrane helix topology with a luminal N-terminus and cytosolic C-terminus were made based on several factors: hydropathy plots, the positive inside rule (from the observation that lysine and arginine residues are more abundant in cytoplasmic regions (von Heijne, 1992)), and the fact that peptide sequences fused to the

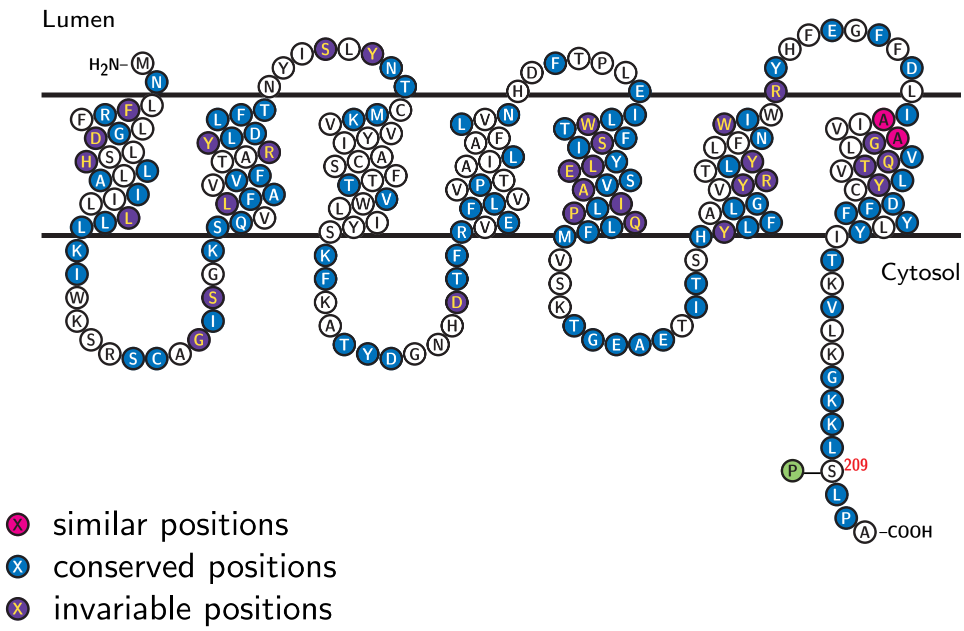


Figure 1.5 Predicted topology diagram of the KDELR. Residue conservation is also shown, based on an alignment of 34 KDELR homologues (Reviewed Swiss-prot sequences).

C-terminus of KDELR can be phosphorylated *in vivo* by cytoplasmic kinases, but cannot be glycosylated by luminal enzymes, suggesting a cytoplasmic C-terminus (Semenza & Pelham, 1992). Further evidence for the now accepted topology, which can be seen in figure 1.5, comes from the fusion of signal peptide to the N-terminal and an additional ER retention signal KK*xx* to the C- terminal (Townsley et al., 1993). A six transmembrane topology based on N-linked glycosylation was proposed by Singh et al. (1993), but this has generally not been accepted by the community.

An extensive mutational analysis of ~40% of the residues in human KDELR1 by Townsley et al. (1993) looked at the localisation and KDEL binding properties in COS cells. Of the many mutations studied, one was of particular interest: D193→N, which showed the ability to bind KDEL *in vitro* but did not relocate to the ER upon stimulation with KDEL tagged lysozyme; rather it remained in the Golgi. D193 is predicted to sit within helix 7, while other mutations of conserved residues in helix 7 appear to have little effect. It has been speculated that this residue may be involved in interactions with other proteins or an oligomerisation event upon cargo binding that would expose a targeting signal for retrograde transport (Townsley et al., 1993).

Building on this work, Scheel & Pelham (1998) set out to identify residues involved in the binding of the KDEL sequence using site-directed mutagenesis and sulfhydryl-specific labelling. They identified several residues potentially involved in ligand binding and showed that R5 and D50 are most likely on the luminal side of the KDELR - providing further evidence for the 7 TM helix topology. D50 was identified as being crucial for H/KDEL but not DDEL binding - potentially showing importance in interactions with a positively charged -4 position of a KDEL motif.

Human KDELR1 and 3 appear to bind a wide range of KDEL motifs, while KDELR2 appears more specialised. KDELR3 favours a HDEL sequence over the canonical KDEL sequence favoured by KDELR1 (Raykhel et al., 2007). Alanen et al. (2011) have shown that the -5 and -6 positions from the C-terminal, that is the two amino acids directly upstream of the KDEL motif, can play a role in modulating the affinity to KDELRs. Some terminal tetrapeptide motifs such as HXEL (X≠D or E) often have well-conserved, charged -5 and -6 residues. The upstream positions appear to have different effects on the three different KDELRs. Mei et al. (2017) explore the possible importance of aromatic residues at the -4 and -6 position of the HDEL motif in yeast and speculate this may be due to $\pi - \pi$ interactions with Erd2. This pattern does not appear in screens of SH-SY5Y cells where secretion of a GLuc fusion protein with a large variety of KDEL motifs was measured (Trychta et al., 2018).

1.2.3 KDEL Receptor Coupling to Vesicle Coat Systems

In order to return ER resident proteins from the Golgi to the ER, the KDELR must engage with cellular trafficking machinery to cycle between these two organelles and fulfil its function. These interactions could be directly to the cytoplasmic vesicle coat machinery, or mediated by other adaptor proteins. In any case, the interactions would have to be selective to the cargo state of the KDELR, such that only cargo-bound KDELR is returned from the Golgi to the ER and, crucially, only empty KDELR returns to the Golgi from the ER.

Little is known about the interactions between the KDELR and its non-cargo interaction partners which mediate anterograde and retrograde vesicular transport as well as other potential partners involved in signalling events. Aoe et al. (1997) reported that KDELR1 interacts with ADP-ribosylation factor GTPase-activating protein 1 (ARF GAP1) and that HeLa cells over-expressing KDELR1 show an ARF1 deactivating phenotype. Cells over-expressing KDELR1 also provided evidence for KDELR1 oligomerisation by immunoprecipitation and blotting with antibodies for two differently tagged KDELR1 constructs being co-expressed. A role for oligomerisation of KDELR in retrograde transport is a hypothesis first proposed by Townsley et al. (1994). ARF GAP1 is involved in COPI recruitment to the membrane and is activated by binding to GTP, Arf1's GTP-GDP cycle is regulated by SEC7-family GEFs and a GAP. It was also shown that activation of KDELR1 by stimulation with lysozyme-KDEL enhances the interaction with ARF GAP1, and this effect is limited by the quantity of KDELR1 expression (Aoe et al., 1998). Förster resonance energy transfer (FRET) experiments in Vero cells have provided further evidence for KDELR oligomerisation upon KDEL ligand binding, from increased FRET signal between two KDELRs expressed with donor and acceptor fluorophores, upon stimulation with KDEL ligand. Subsequent recruitment of ARFGAP was detected by FRET signal between KDELR and ARF GAP1 in a KDEL ligand dependent fashion. The KDELR-ARF GAP1 complex then interacts with a membrane bound Arf1 to recruit COPI (Majoul et al., 2001).

Arakel et al. show a predicted helix immediately C-terminal to the longin domain of δ-COP is required for proper retrieval of HDEL proteins in yeast. Dodonova et al. solved the structure of this helix and suggest its main interaction partner is Arf1, while also being able to make interactions with the membrane. In this model GTP-hydrolysis by Arf1 can modulate the conformation of δ-COP which could be transmitted to binding partners like the KDELR.

KDELR recycling to the ER was may depend on phosphorylation of S209 by protein kinase A (PKA), a S209→A mutation inhibited coatomer/ARF-GAP recruitment thus preventing ER relocalisation and cargo retention. While mutating S209→D could overcome the redistribution phenotype caused by inhibition of PKA on wild type KDELR by mimicking phosphorylation (Cabrera et al., 2003). It is unclear what role PKA phosphorylation plays in the KDELR cycle, perhaps a regulatory one. It may be noted that S209 is not strictly conserved (Cabrera et al., 2003). Furthermore, Townsley et al. (1993) did not observe any relocalisation phenotype in their assay of the S209A mutant.

In a cell line lacking the three ubiquitous Src-like kinases (Src, Yes, Fyn), expressing an activated form of Src caused a relocation of KDELR from the Golgi to the ER, furthermore PE, which uses the KDELR to localise to the ER, was transported at an increased rate without Src, while transport of proteins from the ER to the Golgi appears unaffected (Bard et al., 2003). Src and other Src family kinases were shown to be activated by a KDEL traffic pulse, leading to a phosphorylation cascade that activates intra-Golgi transport machinery (Pulvirenti et al., 2008).

Using a KDEL binding defective mutant KDELR (R169→N), Yamamoto et al. (2003) discovered that the KDELR is involved in ER stress and activation of functional KDELR leads to enhanced phosphorylation of p38 mitogen-activated protein kinases while the binding deficient mutant does not.

The seven predicted transmembrane helices of the KDELR have sparked research to enquire if it may be a distantly related G-protein coupled receptor (GPCR). G-proteins have been observed in the Golgi and are thought to modulate traffic, leading to the hypothesis that the KDELR could potentially function by generating signals in a GPCR-like fashion (Stow et al., 1991; Leyte et al., 1992; Helms et al., 1998; Pimplikar & Simons, 1993; Solis et al., 2017). Some groups have provided evidence that the KDELR activates G-proteins. Giannotta et al. (2012) identify an interaction between KDELR and $G\alpha_q$ which is involved in the activation of Src family kinases (SFKs). Traffic pulses of KDELR ligands appeared to activate $G\alpha_q$, deactivation of this G-protein abrogated SFK activation. Cancino et al. (2014) argue that KDELR activation of $G\alpha_s$ activates PKA which in turn may be involved with Golgi homeostasis along with the $G\alpha_q$ activation pathway. Another G-protein, $G\alpha_o$ is thought to be activated in a monomeric ($G\beta\gamma$ free) state by the KDELR which then activates Rab1/3 signalling which controls outgrowth formation and delivery of material to the plasma membrane (Solis et al., 2017). In all three cases the authors have found that the D193→N mutant identified by Townsley et al. (1993) was unable to activate the G-proteins. These studies indicate that the KDELR may act as a sensor of Golgi traffic and relay this information to a wide variety of cellular processes including membrane trafficking, the actin cytoskeleton and lysosomal repositioning (Tapia et al., 2019).

Our knowledge of the interactions that the KDELR makes, besides the fact that it binds the KDEL peptide motif, has much room to grow. Due to the complexity of the systems that the KDELR is embedded in, direct effects are difficult to disentangle. The nature of the KDELRs interactions on the cytosolic face remain especially enigmatic and further study of these is required for a better understanding of the KDELR and how it communicates with cellular trafficking and signalling complexes.

1.2.4 Evolutionary Relationship of KDELR and PQ-loop Family

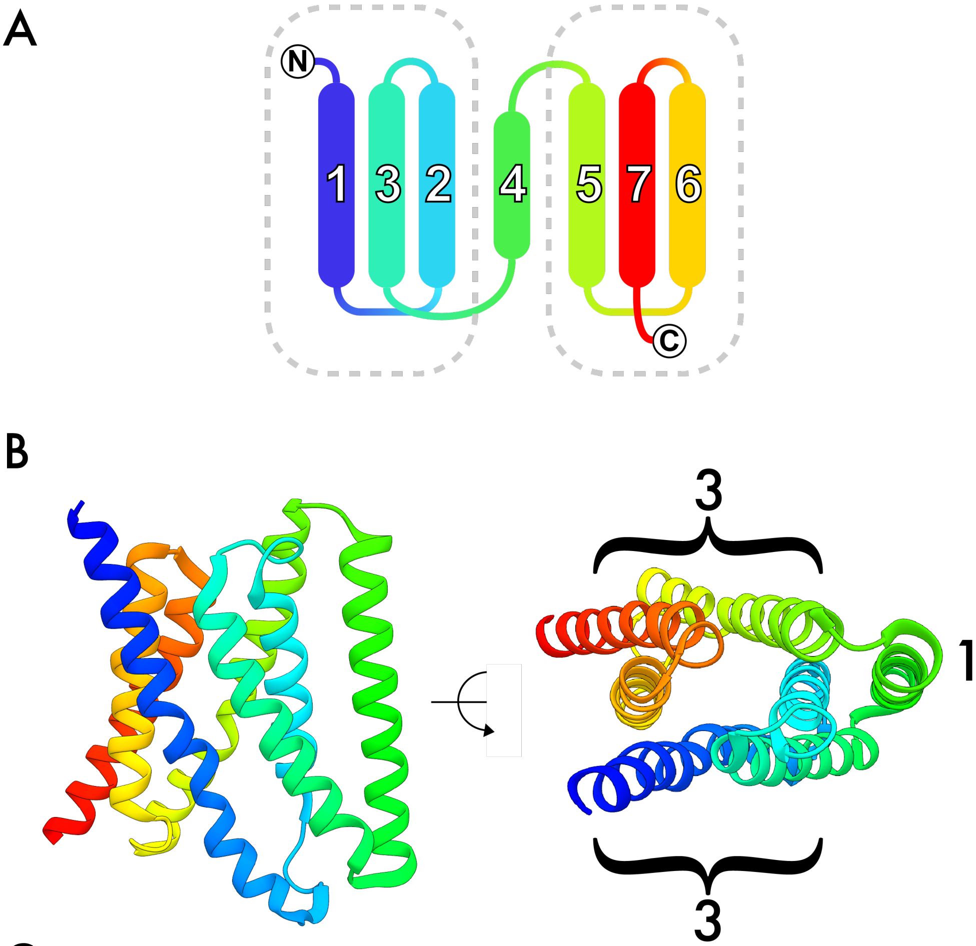
The KDELR forms part of a large and diverse family of membrane proteins called the PQ-loop family (pfam cl21610), which PFAM groups within the MtN3-like clan. PQ-loop family proteins play important roles in lysosomal transport, tissue development and nutrient transport in plants and bacteria, and includes amino acid transporters such as cystinosin (a lysosomal cystine proton symporter) and PQLC2 (a lysosomal cationic amino acid exporter) as well as receptors involved in cellular trafficking, like the KDELR. Many still remain uncharacterised (Feng & Frommer, 2016; Kalatzis et al., 2001; Jézégou et al., 2012; Saudek, 2017). Intriguingly cystinosin has been implicated in both transport and cellular trafficking, displaying both proposed PQ-loop functions (transport and trafficking) (Zhang et al., 2017; Saudek, 2017). Human SWEET1 another PQ-loop protein, in addition to transporting glucose, may have a role in cellular trafficking (Chen et al., 2010; Stokes et al., 2005; Saudek, 2017). Other PQ-loop proteins like Cfs1p, which is involved in phospholipid flippase functions in yeast, complicate this transporter/trafficking picture (Yamamoto et al., 2017).

Several orphan genes with distant homology to the KDELR have also been linked to vesicle trafficking in mammalian cells, suggesting that the KDELR forms part of a larger family important for cellular trafficking. Consistent with this, the general localisation of known PQ-loop members appears to be the membranes of organelles rather than the plasma membrane (Saudek, 2012). It has been postulated that in general the function of PQ-loop proteins is related to protein trafficking, but the nature of this family remains elusive, as does the function of the PQ di-peptide motif from which they derive their name (Saudek, 2012).

From homology searches the KDELR appears to be distantly related to the SWEET sugar transporters, which are members of the PQ-loop family. The structure of *Oryza sativa* SWEET2b, with which human KDELR1 shares 25% sequence identity, is the closest homologous protein for which a structure is known (figure 1.6) (Tao et al., 2015). Structures of bacterial homologues of SWEET transporters (semiSWEETs) have been solved, but these share even less sequence similarity with KDELRs.

The KDELRs and SWEETs are members of the transporter-opsin-G protein-coupled receptor (TOG) superfamily (Yee et al., 2013). The TOG superfamily is a large superfamily of integral membrane proteins and includes a wide variety of proteins including the lysosomal cystine transporter family, the PQ-loop family, nicotinamide ribonucleoside uptake permease family, the GPCR family and several others. Members of the TOG superfamily generally have 7 or 8 transmembrane domains (TMDs), while eukaryotic members of the SWEET family, which appear to be the family most closely related to the KDELR within the TOG superfamily, have 7 TMDs in a 3+1+3 repeat and bacterial members generally have a 6 TMDs in a 3+3 repeat, where the repeats show highest similarity to the first repeat in the 7 TMD proteins Yee et al. (2013). Sequence alignments of KDELRs reveal a similar 3+1+3 pattern.

Are the members of the TOG superfamily derived from a common ancestor with the triple helix bundle (THB) forming the basic building pattern for these proteins (Feng & Frommer, 2016)? Given the KDELRs recognition of a short C-terminal peptide sequence, it is enticing to ask if the KDELR is descended from an ancient transporter, which evolution has effectively co-opted into becoming a receptor, perhaps eliciting reminders of so-called transceptors like Gap1 (a yeast amino acid transporter that also acts as a sensor for amino acid abundance) and SNAT2 (a mammalian amino acid transporter that also senses amino acid abundance) (Donaton et al., 2003; Hyde et al., 2007).



C

```
Os SWEET2b      MDSLYDISCFAAGLAGNIFALALFLSPVTTFKRILKAKSTERFDGLPYLFSLLNCLICLW  60
Hs KDELR1       ---------------MNLF---RFL------------------GDLSHLLAIILLLLKIW  24
                             *:*    **                  ..* :*::::  *: :*

Os SWEET2b      YGLPWVADGRLLVATVNGIGAVFQLAYICLFIFYADSRKTRMKIIGLLVLVVCGFALVSH  120
Hs KDELR1       KSRSCAGISG---KSQVLFAVVFTARYLDLFTNYISLYNTCMKV----VYIACSFTTV--  75
              .   .. .     :   :..**   *: **   * .  :* **:    * :.*.*: *

Os SWEET2b      ASVFFFDQPLRQQFVGAVSMASLISMFASPLAVMGVVI---------------RSESVEF  165
Hs KDELR1       ---WLIYSKFKATYDGNHDTFR-VEFLVVPTAILAFLVNHDFTPLEILWTFSIYLESVAI  131
                ::: . ::  : *  .    :.::. * *::..::                 *** :

Os SWEET2b      MPFYLSLS----------TFLMS----ASFALYGLLLRDFFIYFPNGLGLILGAMQLAL-  210
Hs KDELR1       LPQLFMVSKTGEAETITSHYLFALGVYRTLYLFNWIWRYHFEGFFDLIAIVAGLVQTVLY  191
             :*  : :*           :*::     :: *:. : * .*  * : :.:: * :* .*

Os SWEET2b      ----YAYYSRKWRGQDSSAPLLLA    230
Hs KDELR1       CDFFYLYITKVLKGKKLSLPA---    212
```

Figure 1.6 **A** The 3+1+3 topology of the SWEET2b transporter. **B** The crystal structure of *O. sativa* SWEET2b (PDB ID: 5CTG (Tao et al., 2015)), with the 3+1+3 repeat shown. **C** Sequence alignment of *H. sapiens* KDELR1 with *O. sativa* SWEET2b.

1.3 Aims and Hypothesis

Given the fundamentally important role the KDELR plays in eukaryotic cells there are many unanswered questions regarding how it functions:

- How does the KDELR recognise the KDEL motif, and what is the basis for the pH-dependence of this interaction?
- How does the KDELR interact with its non-cargo binding partners?
- How does the KDELR selectively associate with COPII/COPI vesicles when unbound/bound to cargo and what is the mechanism for this?
- Is the pH gradient between Golgi and ER sufficient to explain the retrograde transport of cargo proteins against their concentration gradient, or is there another driver of this process?
- What is the relationship between the KDELR and other PQ-loop proteins and can we say anything about their evolution?

Answering these questions in a satisfying manner requires many experimental strategies, but most importantly, the atomic structure of the KDELR would enable these to be approached from a more rational starting point. Thus the defining question is:

- What is the structure of the KDELR?

Following a review of the literature surrounding the KDELR the current hypotheses regarding the above questions are:

- C-terminal peptides are likely recognised with a large variety of side chain interactions. Given previous mutational work by **Townsley et al.** and **Scheel & Pelham**, there are several residues that are likely to be involved in peptide recognition, including R5, D50, Y162, and N165. Furthermore, it seems likely that there are specific spacial restrictions or interactions with the C-terminal Leucine residue of the KDEL motif, as this appears to be the most strongly conserved position. Finally, as the interaction is pH-dependent, it is conceivable that a protonation event of a side-chain causes an affinity switch between the Golgi and ER, given the pH range, histidine would be the most obvious candidate.

- There is much less evidence regarding interactions on the luminal side of the KDELR, the most well characterised mutation predicted to be on this side is the D193→N mutation which retains the ability to bind the KDELR motif but does not engage in retrograde transport. The C-terminal part of the KDELR also has a motif reminiscent of the KK*xx* and K*x*K*xx* COPI retrieval sequence, though it is not at its extreme C-terminus. Predicted cytosolic loop five also contains a potential di-acidic COPII binding motif, though selection for ER export is often a more complex procedure (**Nishimura & Balch**, 1997; **Sato & Nakano**, 2007). One might thus expect 3D recognition epitope, rather than 2D sequence motifs for COPI/II recruitment.

- There is evidence that binding of a KDEL ligand to the KDELR activates its retrieval from the Golgi, so it stands to reason that the KDELR communicates its bound and unbound state to its cytosolic state. In this way it may display or hide interaction sites for COPII and COPI machinery. If and how this is achieved is still largely a mystery.

- A pH gradient is the only currently suggested driver of the affinity switch between the KDELR and its ligands. If the difference in affinity between low and high pH states is sufficient and is communicated across the membrane to the cellular trafficking machinery, the pH difference between ER and Golgi may be sufficient to drive this sorting process, even against its concentration gradient.

- Given the bioinformatic analysis it would appear that the KDELR will share the 3-1-3 topology of the PQ-loop family, with a similar THB architecture.

Structural insights may shed light on the question regarding its status as a broken transporter turned receptor.

To explore these questions this thesis will present several crystal structures of the KDELR: in the apo state, bound to KDEL/HDEL peptides, and bound to an antagonistic synthetic nanobody (Sybody). In addition to the structural data, KDEL peptide binding assays were developed to investigate peptide binding and pH-dependence. In collaboration with Andreas Gerondopoulos, *in vivo* cell based assays were also performed to validate structural and *in vitro* findings in a cellular context.

2 Screening & Characterisation

2.1 Homologue Screening

In order to maximise the chances of a successful crystallographic project, it is imperative to use as many means as reasonably possible to turn the odds in ones favour. One of the ways this is achieved is by screening a range of homologues of the protein of interest to identify those which possess the most desirable characteristics for crystallography. These include: expression in sufficient quantity for purification, amenability to purification in high yield and purity, and thermal stability. The first two characteristics are desirable to facilitate the high throughput of screening crystallisation conditions while the latter two are considered properties that will increase the likelihood of crystallisation, as a more stable and thus presumably more rigid molecule should be able to form and maintain crystal contacts more readily than a flexible and unstable one. As it is not possible to determine these qualities *a priori* from the amino acid sequence of a protein, it is necessary to perform experimental screening to determine them.

As the KDELR is strongly conserved in eukaryotes, a wide selection of homologues is available. In this case several targets, see table 2.1, were chosen for different reasons, including: the source organism is of interest as a model for this system (e.g. *Saccharomyces cerevisiae* (*Sc*) & *Homo sapiens* (*Hs*)), the source organism is a thermophile (*Galdieria sulphuraria* (*Gs*) & *Chaetomium thermophilum* (*Ct*)) and thus its proteins may demonstrate increased thermostability, and to cover a wide range of different organisms in order to capture this variety in the screening process (e.g. *Gallus gallus* (*Gg*) & *Trypanosoma brucei* (*Tb*)). A sequence identity of over 20% is often sufficient to inform structural homology (**Wilson et al., 2000**).

Organism	Uniprot	Gene Name	Motif	Length	Mass (Da)
Homo sapiens	P24390	KDELR1	KDEL	212	24,542
Mus musculus	Q99JH8	KDELR1	KDEL	212	24,560
Rattus norvegicus	Q569A6	KDELR1	KDEL	212	24,546
Gallus gallus	Q5ZKX9	KDELR2	KDEL	212	24,456
Caenorhabditis elegans	P48583	ERD2	KDEL	213	25,060
Arabidopsis thaliana	P35402	ERD2	HDEL	215	25,204
Oryza sativa	Q65XC2	Os05g0529300	HDEL	215	25,520
Saccharomyces cerevisiae	P18414	ERD2	HDEL	219	25,763
Trypanosoma brucei	Q384K5	Tb11.01.0410	MDDL/KQDL	219	25,956
Galdieria sulphuraria	M2XM93	Gasu_15420		219	25,814
Chaetomium thermophilum	G0S1J0	CTHT_0013770	HDEL	209	24,574

Table 2.1 KDELR homologues screened during the course of the project.

Solving the structure of the KDELR of one organism should therefore provide a rich source of information regarding the structure of all the KDELRs as they are very strongly conserved. As on can see in table 2.2 and figure 2.1, the KDELRs of varying organisms share a high degree of similarity, with identities over 40%. This indicates the function of the KDELR is likely to play an important role in eukaryotic organisms.

2.1.1 Expression and FSEC

Eleven homologues (table 2.1) of the KDELR were cloned into a pDDGFP2-Leu2d vector for expression in *S. cerevisiae* (see methods). These constructs are thus all N-KDELR-TEV-GFP-8xHis-C. All of the constructs, except *Sc* KDELR showed appreciable levels of expression as determined by Green Fluorescent Protein (GFP) fluorescence measured at the time of harvest. Given the satisfying results, all expressing homologues were taken forward to the next stage of screening.

	H. sapiens	*M. musculus*	*R. norvegicus*	*G. gallus*	*C. Elegans*	*A. thaliana*	*O. sativa*	*S. cerevisiae*	*T. brucei*	*G. sulphuraria*	*C. thermophilum*
H. sapiens	—	100.0	100.0	93.3	83.9	69.8	66.5	69.1	66.3	69.7	71.0
M. musculus	99.5	—	100.0	93.3	83.9	69.8	66.5	69.1	66.3	69.7	71.0
R. norvegicus	99.0	99.5	—	93.3	83.9	69.8	66.5	69.1	66.3	69.7	71.0
G. gallus	84.4	83.9	83.9	—	83.0	70.8	66.9	68.7	66.3	72.1	72.0
C. Elegans	65.5	65.0	64.6	65.0	—	71.9	68.5	68.3	66.9	68.8	71.2
A. thaliana	50.7	50.7	51.1	50.2	48.0	—	86.9	63.5	64.3	74.6	67.6
O. sativa	51.1	51.1	51.6	47.8	47.6	76.7	—	63.0	61.9	73.2	67.6
S. cerevisiae	49.7	49.2	49.2	50.2	44.8	48.1	44.3	—	58.4	65.4	71.8
T. brucei	48.8	48.3	48.8	50.7	49.5	45.5	44.1	39.2	—	66.9	64.7
G. sulphuraria	51.9	51.4	51.9	51.4	46.4	53.5	53.5	48.1	45.7	—	69.6
C. thermophilum	54.9	54.4	53.9	56.8	51.7	49.5	50.0	54.8	44.1	50.4	—

% identity

% similarity

Table 2.2 Sequence identity and similarity of the screened homologues.

Detergent Solubilisation

An important factor in membrane protein purification is the choice of detergent. Different membrane proteins require different detergent conditions to remain stable outside of the membrane. There are many variables that affect detergent interactions with proteins, from the chain length to the head group chemistry. In general the more charged a detergent head group is, the harsher it is, similarly the smaller the head group the harsher (Privé, 2007). Finally the shorter the alkyl tail of the detergent the harsher it will tend to be. Harshness here refers to the likelihood of the detergent in question denaturing the protein being solubilised (Privé, 2007). To find an optimal detergent for the crystallisation of a target protein, screening is required as which will be effective is not known beforehand (Keyes et al., 2003). An ideal detergent should fulfil several criteria: firstly it should solubilise the target protein from the membrane efficiently, secondly the solubilised protein should remain stable and retain its secondary structure once solubilised, and thirdly the detergent should be amenable to crystallographic techniques.

The most commonly used detergents in alpha helical membrane protein crystal structures are maltosides (50%, of alpha helical structures in the PDB), glucosides (23%) and amine oxides (7%) (Parker & Newstead, 2016). It has been

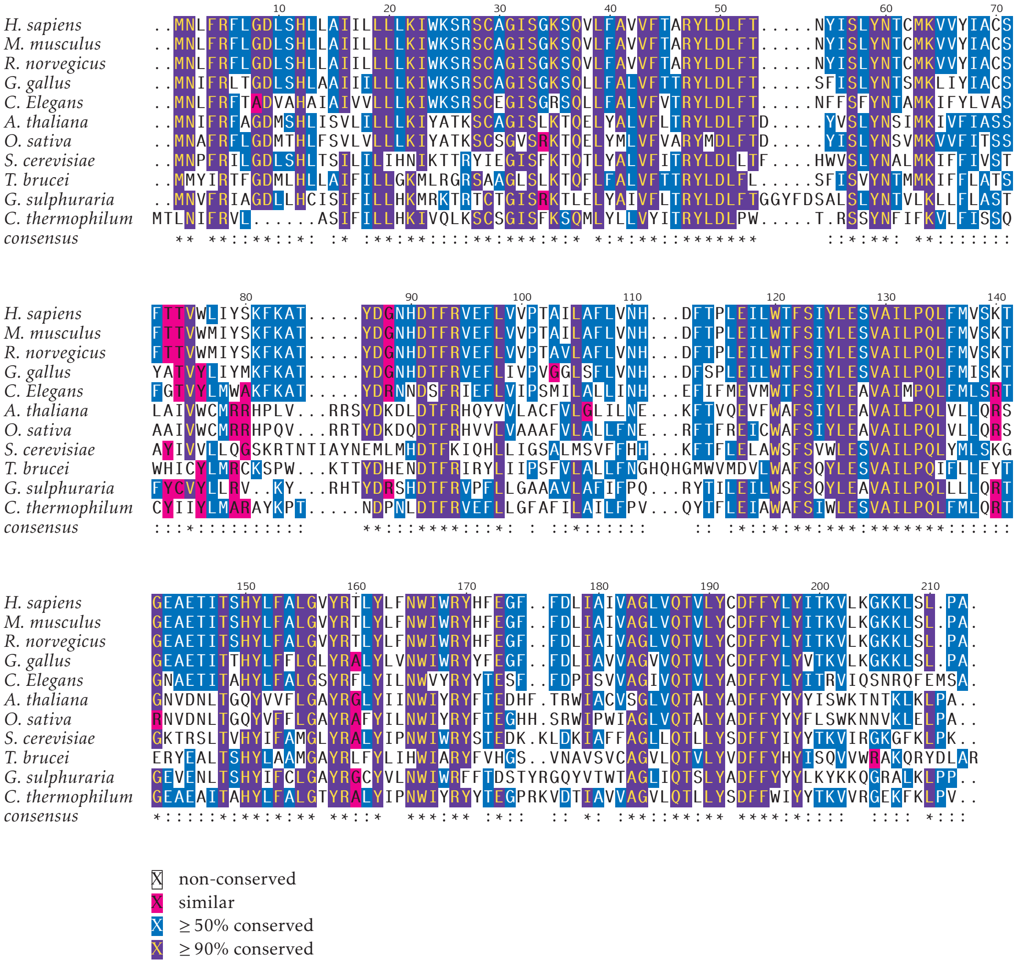


Figure 2.1 Sequence alignment of the amino acid sequences of all the KDELRs screened. Numbering is according to the human KDELR.

observed that shorter chain detergents can be more conducive to higher resolution structures as longer acyl chains encompass more of the protein thus preventing crystal contacts, though short chain detergents are generally hasher (Parker & Newstead, 2016). Membrane proteins in vapour diffusion crystals must make crystal contacts while surrounded by detergent micelles, the choice of which can affect possible crystal packing and contacts, Lipidic Cubic Phase (LCP) replaces the detergent with a lipid so detergent choice has less of an impact on crystallisation in this instance. To screen detergents efficiently a sparse screen was performed, while testing DDM with all of the homologues, to allow for easier comparison. As membrane protein stability is predominantly intrinsic and thus detergent independent, screening a variety of different detergents should give an idea of the stability of the homologues as well as the suitability

Detergent	Class	Chain Length	CMC	Micelle Size[1]
DDM	Non-ionic Maltoside	12	0.17 mM	72 kDa
DM	Non-ionic Maltoside	10	1.8 mM	40 kDa
NM	Non-ionic Maltoside	9	6 mM	
LMNG	Non-ionic Maltose-Neopentyl Glycol	12	0.01 mM	91 kDa
DMNG	Non-ionic Maltose-Neopentyl Glycol	10	0.036 mM	
LDAO	Zwitterionic Amine Oxide	12	1–2 mM	21.5 kDa
NG	Non-ionic Glucoside	9	6.5 mM	90 kDa
DHPC	Phosphocholine lipid	7	1.4 mM	
CHAPS	Zwitterionic	N/A[2]	8 mM	6 kDa
CYMAL-5	Non-ionic Maltoside	N/A[2]	2.5 mM	23 kDa

Table 2.3 See Acronyms and Abbreviations section for full names. [1] Data from anatrace website if available or (Stetsenko & Guskov, 2017), blank if no source found [2] These detergents do not have simple alkyl chains. Cymal detergents have a cyclohexyl aliphatic tail while CHAPS is derived from bile salts and has a cholesterol-like structure.

of the various detergents tested (Sonoda et al., 2011). Furthermore cholesteryl hemisuccinate (CHS) has been observed to stabilise some membrane proteins, even independently of any physiological relevance to the presence of CHS, thus a DDM-CHS and DM-CHS mixtures were also tested, see table 2.3 (Thompson et al., 2011; Jungnickel et al., 2018).

As figure 2.2 demonstrates, solubilisation efficiency alone does not provide a lot of information about the stability of different homologues and the values are general rather than concrete. It is possible to rule out several detergents, such as NG or CHAPS based on this screening process, as they solubilise less than half of the protein present. To obtain more information about the behaviour of the different homologues a more sophisticated approach is required.

Fluorescence Size Exclusion

Fluorescence Size Exclusion Chromatography (FSEC) provides a fast and high throughput method for screening membrane proteins (Kawate & Gouaux, 2006; Newstead et al., 2007). As all of the constructs are tagged C-terminally with a GFP, a crude solubilised membrane fraction may be analysed on a Size Ex-

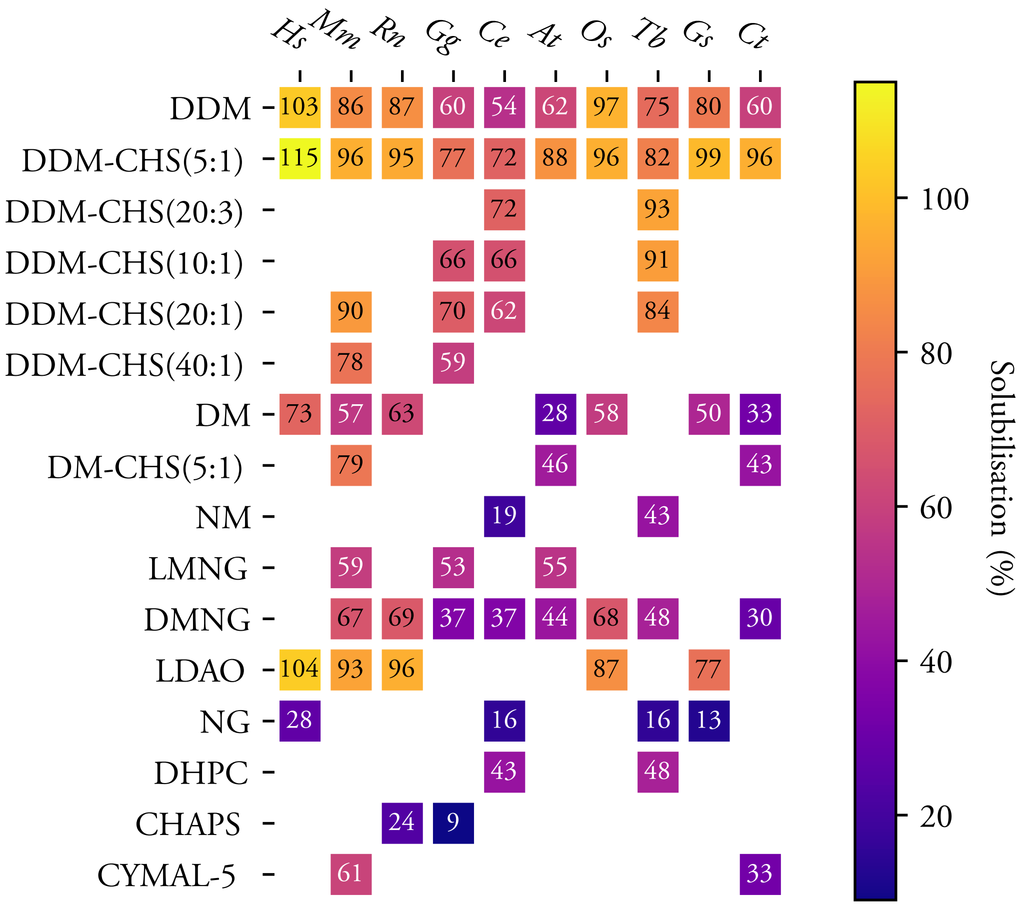


Figure 2.2 Detergent solubilisation efficiencies of various detergents with different homologues. These values are rough estimates as solubilisation may vary from batch to batch. Ratios of CHS are in w/w. Solubilisation efficiency is calculated by measuring of GFP fluorescence of solubilisation mixture before ultracentrifugation and the supernatant after ultracentrifugation.

clusion Chromatography (SEC) column and monitored at the GFP fluorescence (Excitation 488 nm, Emission 512 nm). In this way the protein of interest may be distinguished from the other proteins in the fraction and the quality of the protein may be assessed, based on the monodispersity of the observed protein species (Kawate & Gouaux, 2006).

Figure 2.3 shows a selection of FSEC traces for all of the KDELR constructs. From this FSEC screen the two thermophilic candidates (*Gs* and *Ct*) were excluded due to their poor behaviour, with a large quantity of the peaks being in the void volume. An interesting trend that was observed in all the FSECs was that DDM-CHS mixtures appeared to produce the largest, monodispersed peaks. Thus for initial small scale purifications DDM-CHS (5:1 (w/w)) was chosen, as this mixture appears to provide excellent solubilisation and does not appear to denature the KDELR or cause significant aggregation. However, as favourable behaviour in an FSEC does not necessarily translate into favourable behaviour during purification, a small scale purification was performed on all of the homologues.

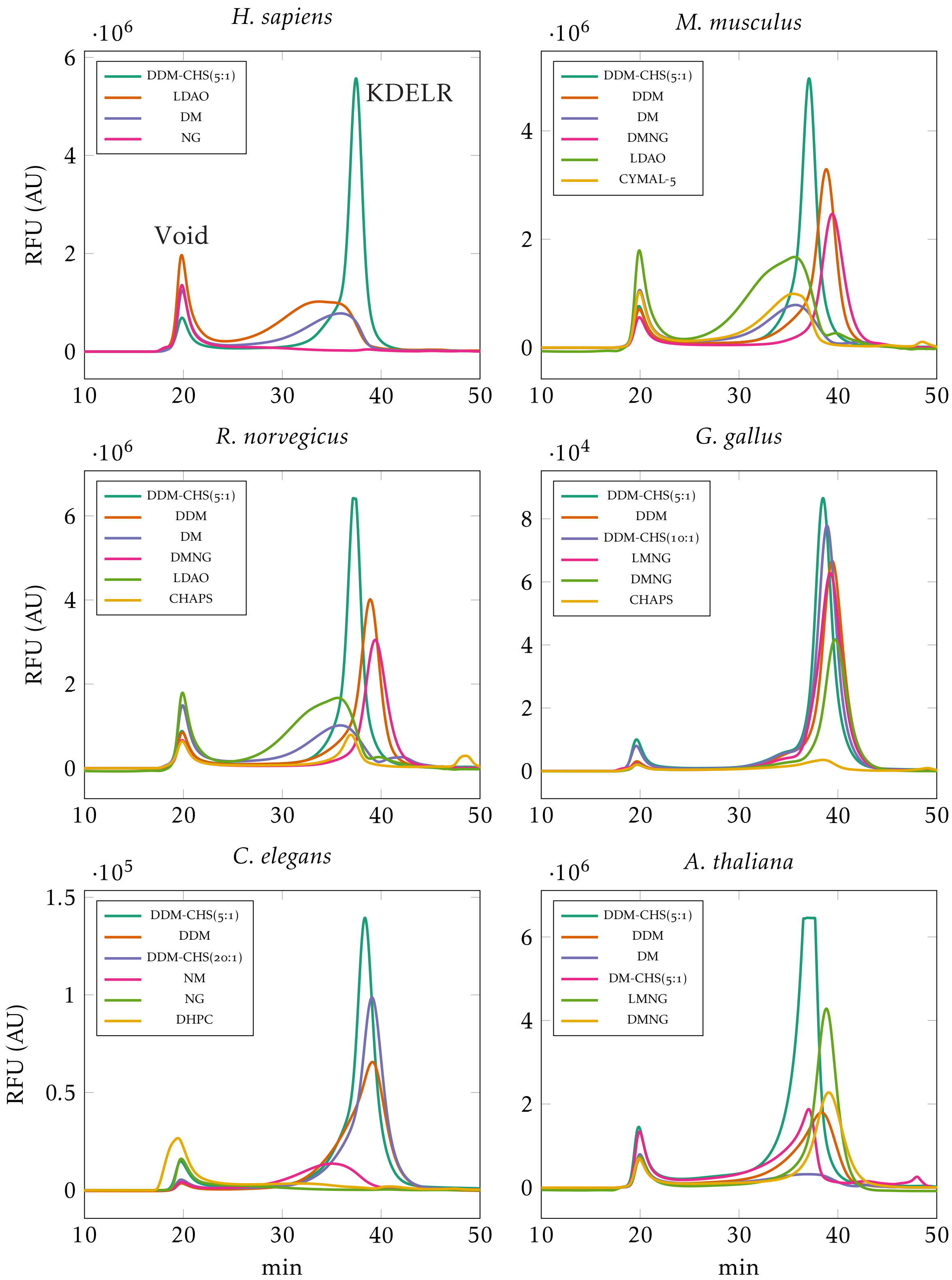


Figure 2.3 FSEC traces of the homologues tested with various detergents, run on a Superose 6 SEC column. No molecular weight standards were used.

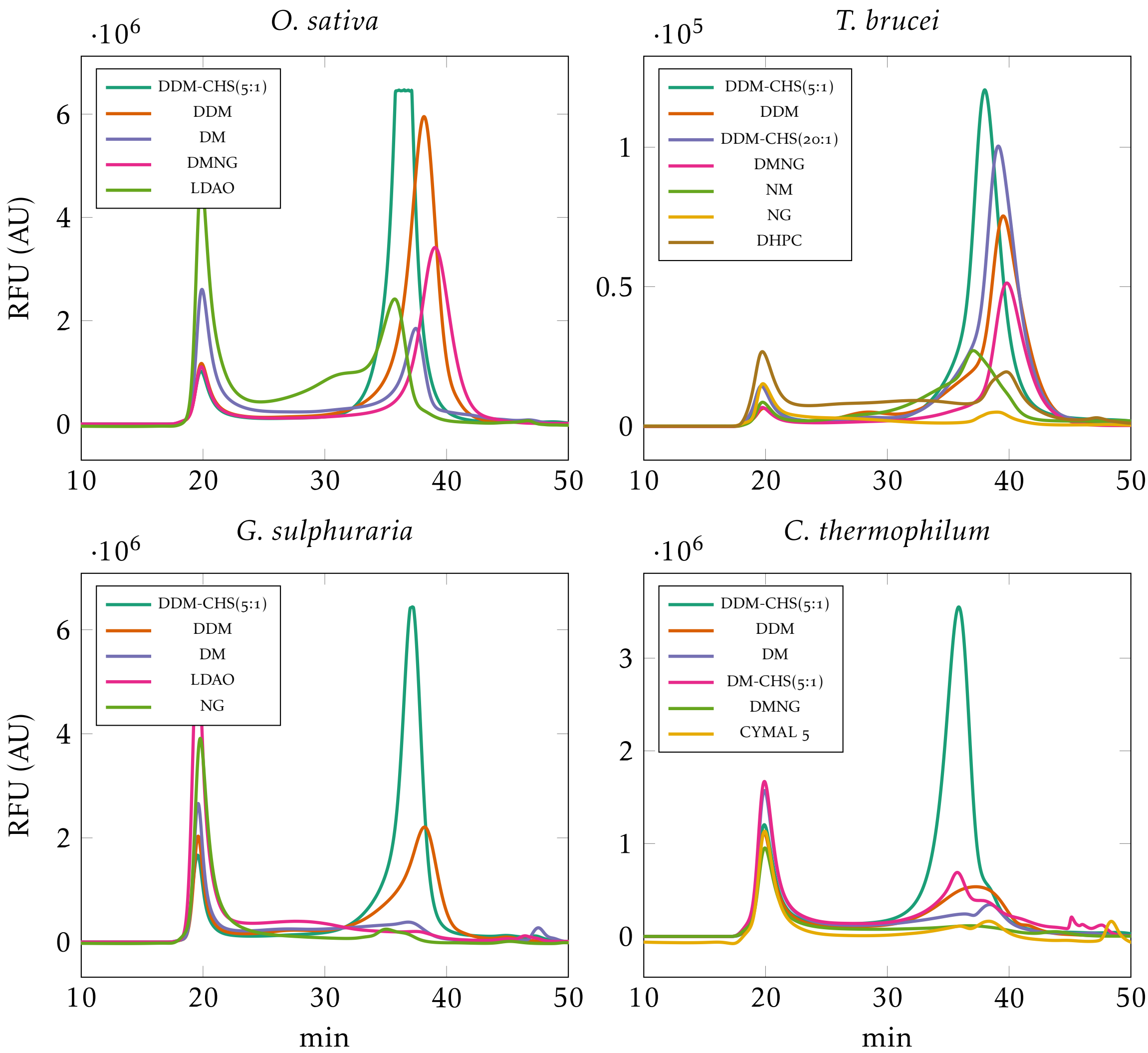


Figure 2.3 (Continued) FSEC traces of the homologues tested with various detergents, run on a Superose 6 SEC column. No molecular weight standards were used.

2.1.2 Small Scale Purification

The small scale purification protocol from 2–5 g of cell membranes from 3 L of yeast culture, consists of an Ni-affinity step, followed by overnight dialysis with Tobacco Etch Virus (TEV) protease, cleaving off the GFP and the C-terminal His-tag. Then a reverse Immobilised Metal Affinity Chromatography (IMAC) is performed, the flow through is concentrated and subsequently loaded onto an S200 column. The resulting S200 traces are shown in figure 2.4, with SDS-PAGE gels of the *Mus Musculus* (*Mm*), *Gg* and *Caenorhabditis elegans* (*Ce*) KDELR purifications shown in figure 2.5.

The effect of CHS

It has been observed that CHS added to DDM micelles alters their architecture by forming a bicelle-like structure, with a less curved centre region and the addition of CHS to DDM micelles has in several cases, been observed to increase the thermal stability of solubilised proteins (**Thompson et al.**, 2011). The addition of increasing amounts of CHS to DDM micelles has been reported to increase the size of these, which we were able to confirm with Size Exclusion Chromatography - Multi-Angle Light Scattering (**Thompson et al.**, 2011). In figure 2.6 we observe that a decreasing amount of CHS added to a DDM micelle decreases the size, indicated by the shift of the peak to the right and the corresponding decrease in molar mass calculated from the multi-angle light scattering. The larger size and modified shape of these CHS containing micelles may offer a more protective environment to solubilised membrane proteins, explaining the increase in thermal stability (**Thompson et al.**, 2011). Others have observed that in molecular dynamics simulations of GPCRs the presence of CHS in the micelle increased the stability of α-helices with direct interactions between CHS and the helices (**Lee et al.**, 2016).

FSEC experiments and small scale purifications showed that the ratio of DDM:CHS had an effect on purification of KDELR homologues. While DDM showed promise in FSEC traces, purification of KDELR appears to require the presence of CHS. In figure 2.7 the effect of DDM:CHS ratios on the purification of *Gg* and *Ce* KDELR is shown. Neither could be successfully purified in the absence of CHS, precipitating during overnight dialysis. Increasing amounts of CHS shifted the SEC peaks to the left, indicating a larger size. Intriguingly two distinct peaks appear in the *Gg* KDELR SEC traces, perhaps a dimeric and monomeric peak. Given the larger size of higher CHS micelles, dimerisation may be stabilised by these, while the monomeric form is predominant in lower CHS micelles. However, SEC-MALS measurements (figure 2.8) of *Gg* and *Mm* at DDM-CHS (20:1) and (10:1) indicate monomeric proteins of ~27 kDa and ~24 kDa respectively.

Several related methods all demonstrated an increasing thermal stability of various KDELR homologues with increasing CHS content in the micelle. These include Circular Dichroism (CD) (no full series of CHS concentrations was measured so data is not shown), nano differential scanning fluorimetry (nanoDSF) (see figure B.2 for *Mm* KDELR melts with increasing CHS concentrations). The latter of these methods monitors intrinsic tryptophan fluorescence of the proteins in response to a temperature gradient, while CD measures the difference in absorption of circularly polarised light. The presence of secondary structure elements in a protein produce characteristic spectra and the signal of these spectra are then monitored over a temperature gradient. Thus sensitivity to CHS concentration provided another parameter with which to distinguish between homologues. A tolerance for lower CHS concentrations was preferred for three reasons. Firstly, purifications with less CHS appeared to result in higher purity. Secondly, smaller micelle sizes are generally desirable for crystallography. And finally, ability to be purified in the presence of less CHS may be seen as an indicator of greater intrinsic stability in this case.

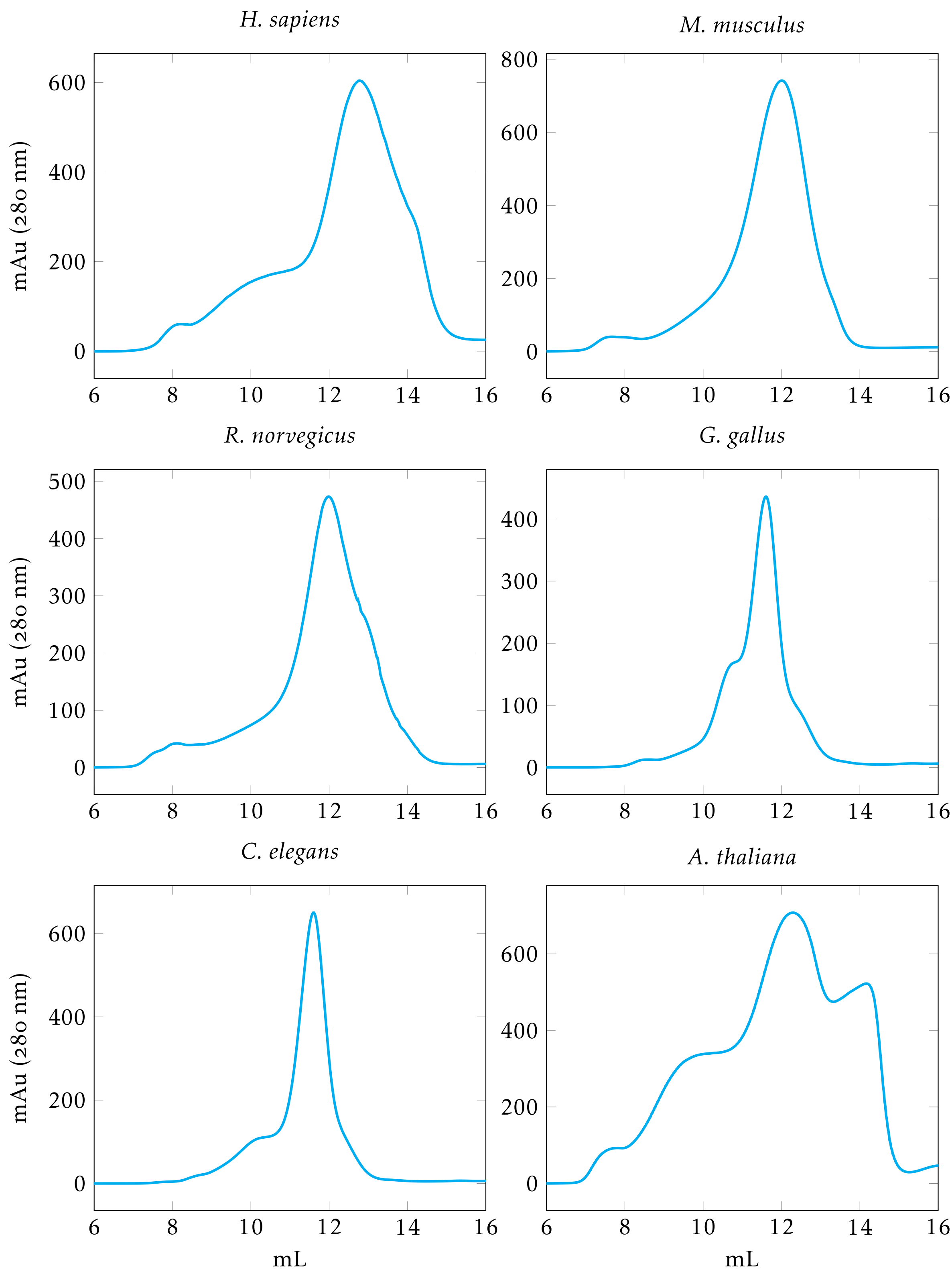


Figure 2.4 SEC traces of the final run on an S200 column of all the homologues purified on a small scale, in DDM-CHS (5:1). Columns were not calibrated with molecular weight standards.

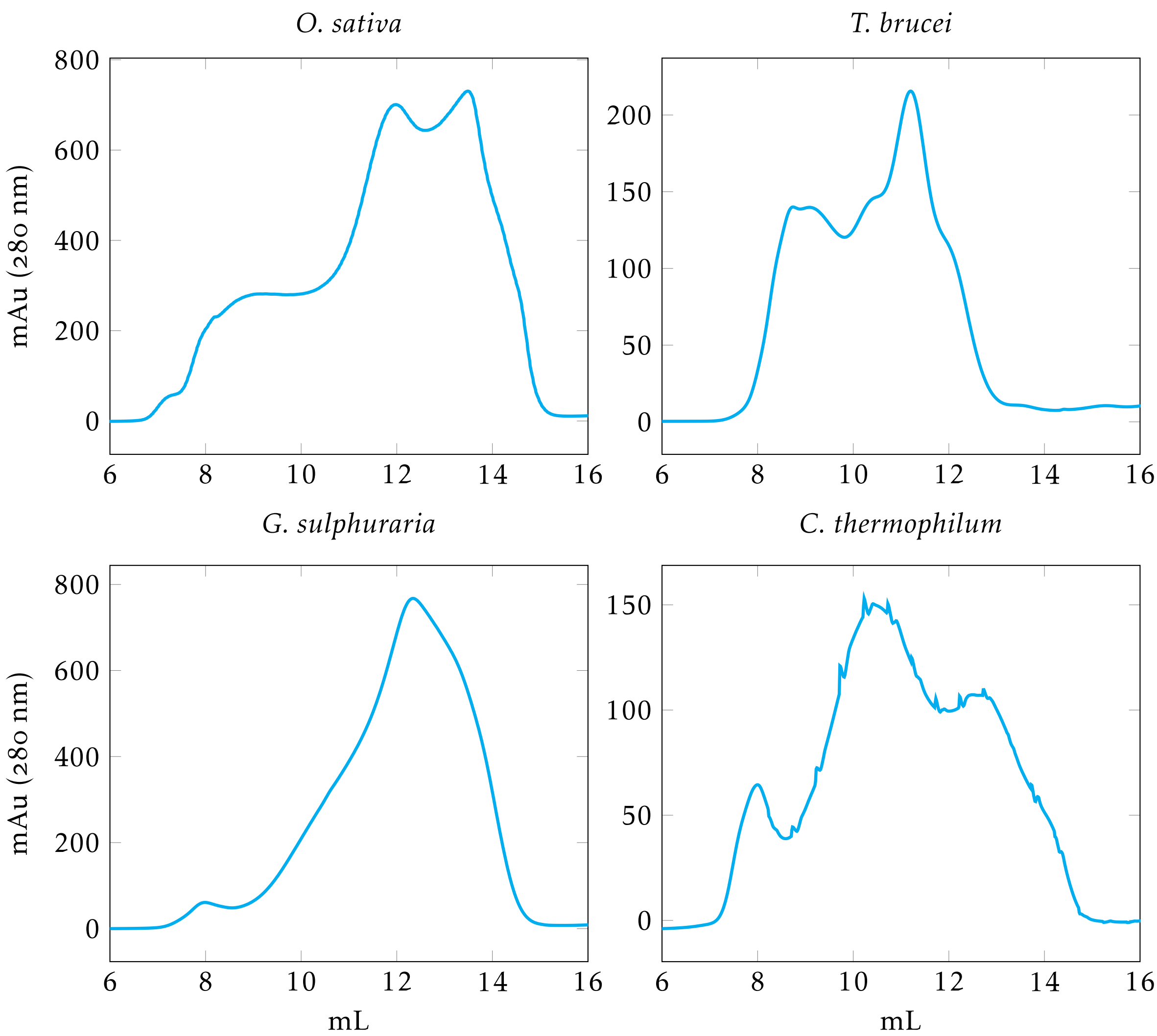


Figure 2.4 (Continued) SEC traces of the final run on an S200 column of all the homologues purified on a small scale, in DDM-CHS (5:1). Columns were not calibrated with molecular weight standards.

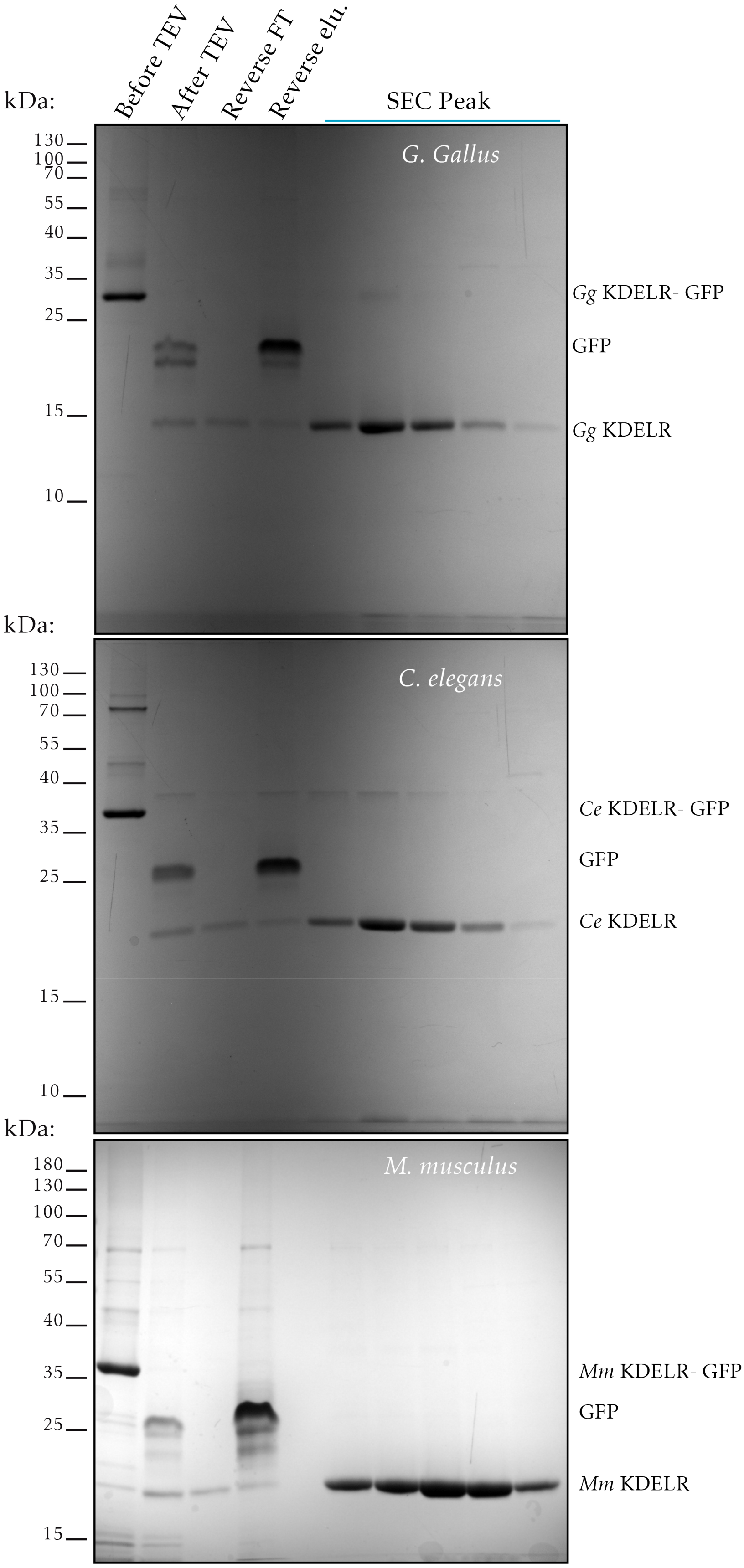


Figure 2.5 SDS-PAGE of *Gg*, *Ce* and *Mm* KDELR small scale purifications. Stained with coomassie brilliant blue.

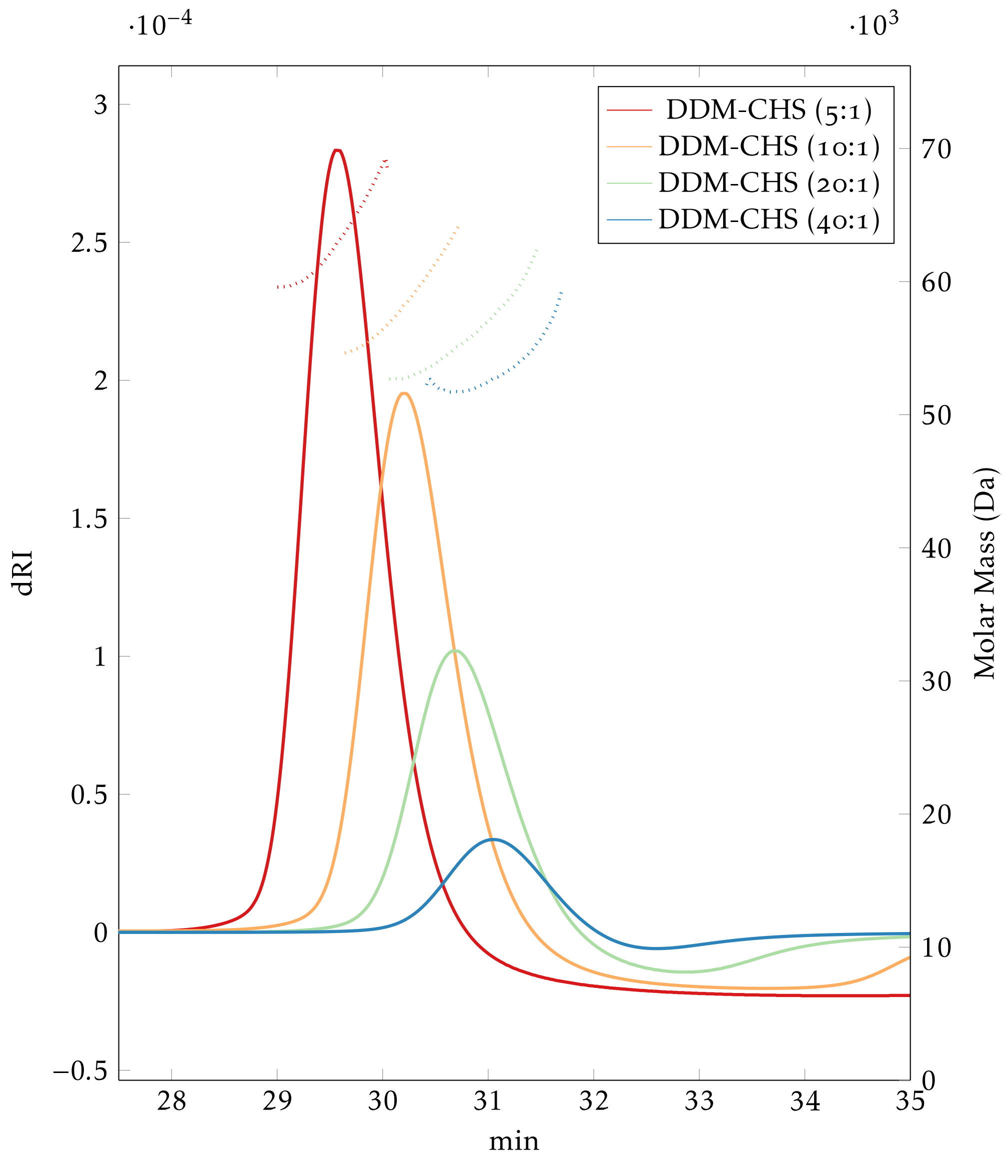


Figure 2.6 SEC-MALS traces of empty DDM-CHS micelles with varying ratios of CHS to DDM (w/w). Dashed lines indicate the calculated molar mass of the empty micelles DDM-CHS(5:1) 62 kDa , DDM-CHS(10:1) 58 kDa, DDM-CHS(20:1) 56 kDa, DDM-CHS(40:1) 53 kDa.

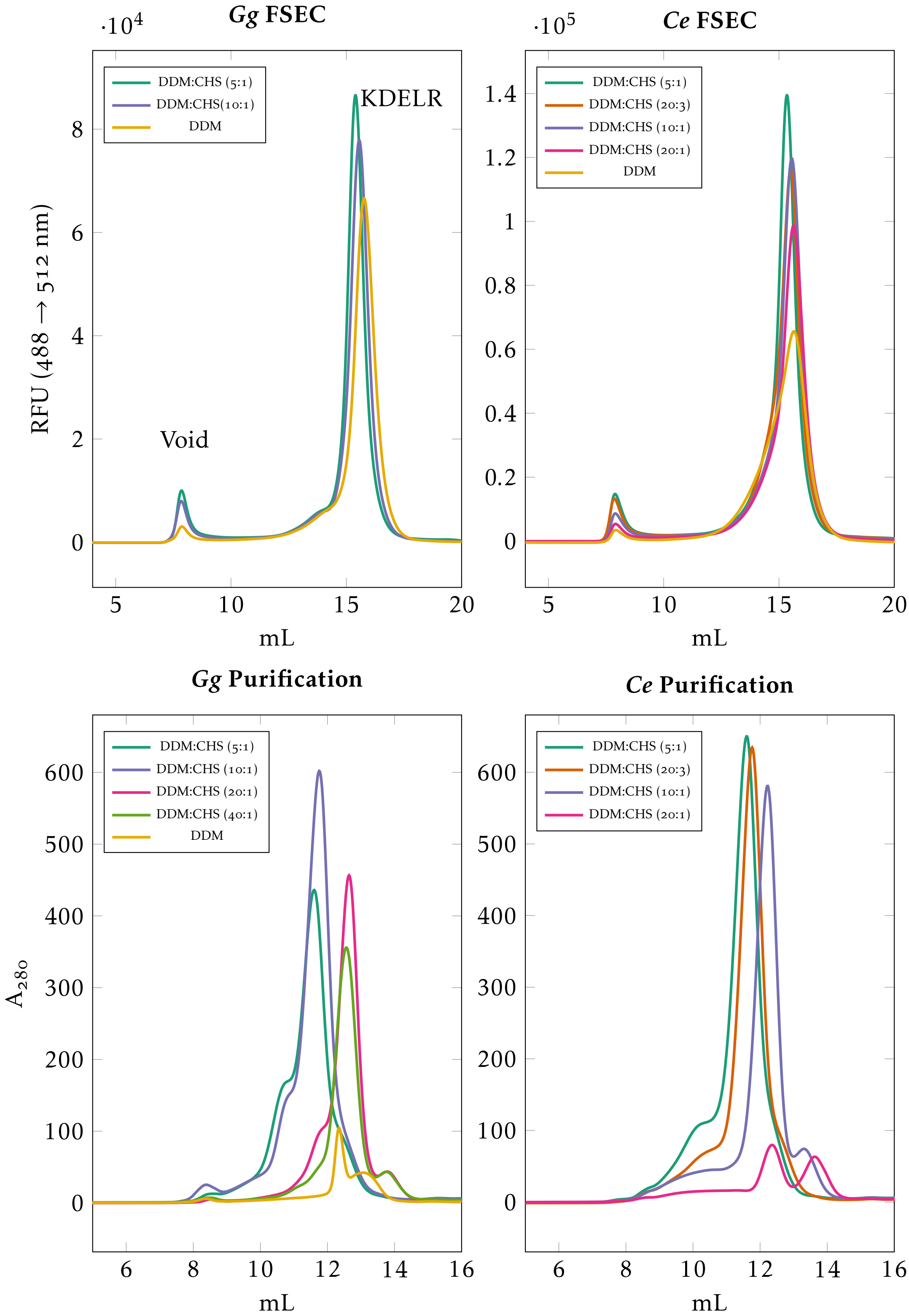


Figure 2.7 *Gg* and *Ce* FSEC (Superose 6 column) and purification (S200 column) with different ratios of DDM:CHS. Columns were not calibrated with molecular weight standards.

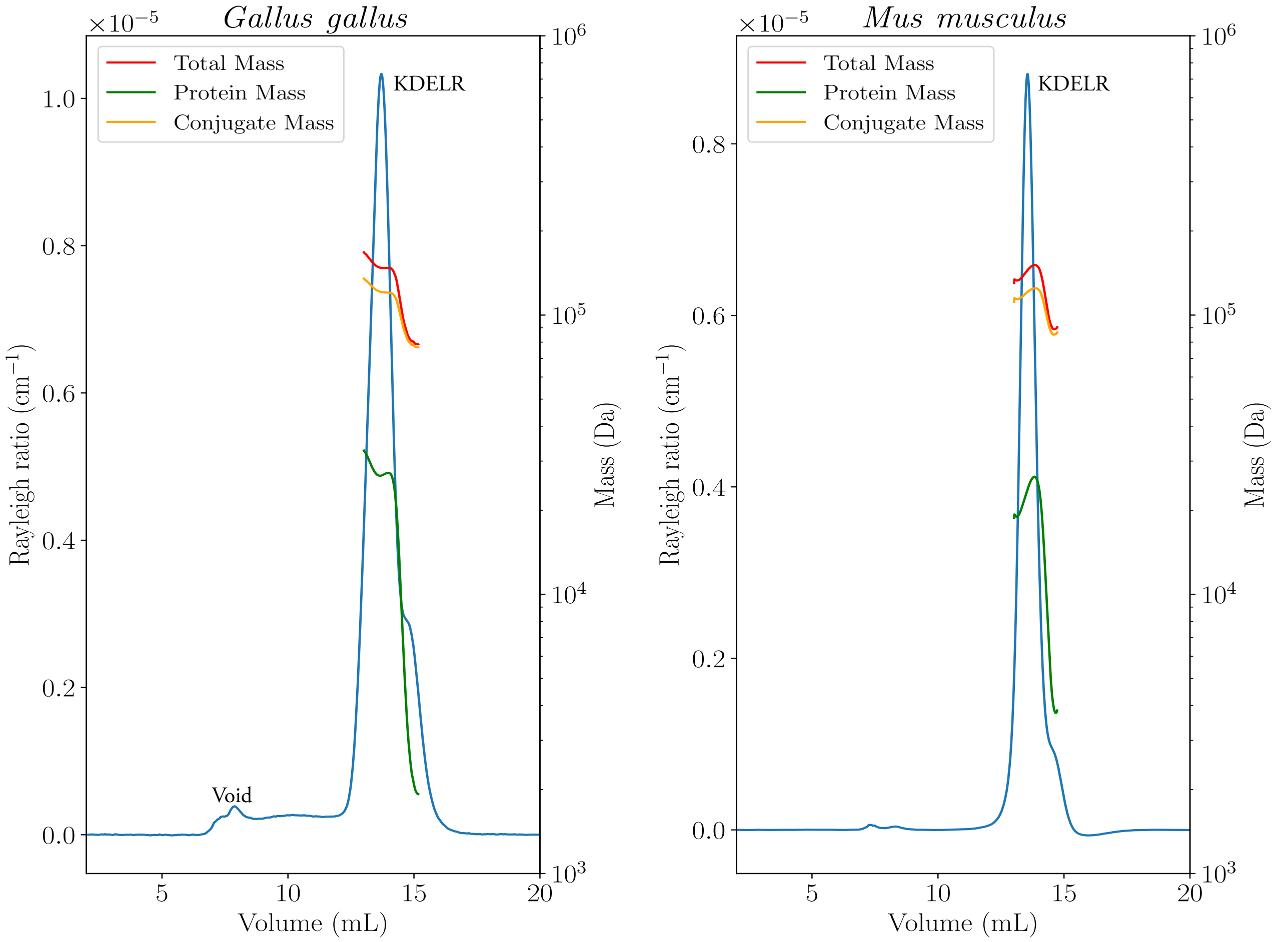


Figure 2.8 SEC-MALS of *Gg* in DDM-CHS (20:1) and *Mm* in DDM-CHS (10:1). At the centre of the peak, the estimated mass for *Gg* is 27 kDa and 24 kDa for *Mm*, corresponding to a monomer peak. Both display a detergent shoulder to the right of the peak, which appears to consist mainly of empty micelles.

Organism	Expression	Solubility	FSEC	Small Scale	Purity	DDM:CHS
Hs	Yes	+++	+	+	+++	
Mm	Yes	+++	++	+	+++	10:1
Rn	Yes	+++	++	+	+++	
Gg	Yes	++	+++	++	+++	40:1
Ce	Yes	++	++	++	+++	10:1
At	Yes	+++	++	-	++	
Os	Yes	+++	++	-	++	20:3
Sc	No					
Tb	Yes	++	+++	--		
Gs	Yes	+++	-	--		
Ct	Yes	+++	-	--		

Table 2.4 KDELR homologues screened during the course of the project showing various properties. Classifications are based on the relative performance within the group of homologues. Expression was judged from GFP fluorescence counts. Solubility on the GFP fluorescence measurements taken before and after detergent solubilisation + > 50%, ++ > 75%, +++ > 90%. FSECs were judged based on apparent monodispersity in a variety of detergents, - mostly void peaks, + one or fewer monodispersed protein peak, large void peaks, ++ monodispersed protein peaks in three or more detergents with little void, +++ monodispersed protein in most detergents with minimal void peaks. Small scale purifications were judged on their SEC trace profile, -- failed to purify due to precipitation, - no monodispersed protein in SEC, + monodispersed protein, ++ monodispersed protein with minimal width. Purity was determined by SDS-PAGE after small scale purification. +++ > 90% ++ > 80%. DDM:CHS refers to the minimum amount of CHS required to successfully purify the protein. Blank spaces indicate the experiment was not performed.

2.1.3 Thermal Stability

From the screening performed on the KDELR homologues, three appeared to have the most desirable qualities for crystallographic work (*Mm*, *Gg* and *Ce*), see table 2.4. The thermal stability of these homologues was tested by CD melts and nanoDSF, table 2.5 shows that *Gg* appeared to be the most stable, followed by *Mm* and *Ce*. The CD melts in figure 2.9 cannot be easily compared to the nanoDSF data as they were measured in different buffers, with *Ce* also having more CHS present, though the differences between *Gg* and *Mm* appear much more pronounced in the nanoDSF measurements. CD measurements in figure 2.9 confirmed that secondary structure was present in all three. As predicted from sequence analysis, the CD data indicated alpha helical secondary structure for all three, with significant negative bands near 222 and 208 nm and a positive band towards 193 nm (**Holzwarth & Doty**, 1965; **Greenfield**, 2006).

	Inflection Point (°C)		
pH	*G. Gallus*	*M. musculus*	*C. elegans*
pH 5	50.0	37.9	22.9
pH 6	52.4	42.9	27.3
pH 7	51.0	43.8	33.0
pH 8	46.8	39.5	34.0

Table 2.5 Inflection points of *Gg*, *Mm*, and *Ce* KDELRs, as determined by nanoDSF. All measurements were in 10 mM potassium phosphate, 50 mM NaCl, 0.03% (20:1) DDM:CHS.

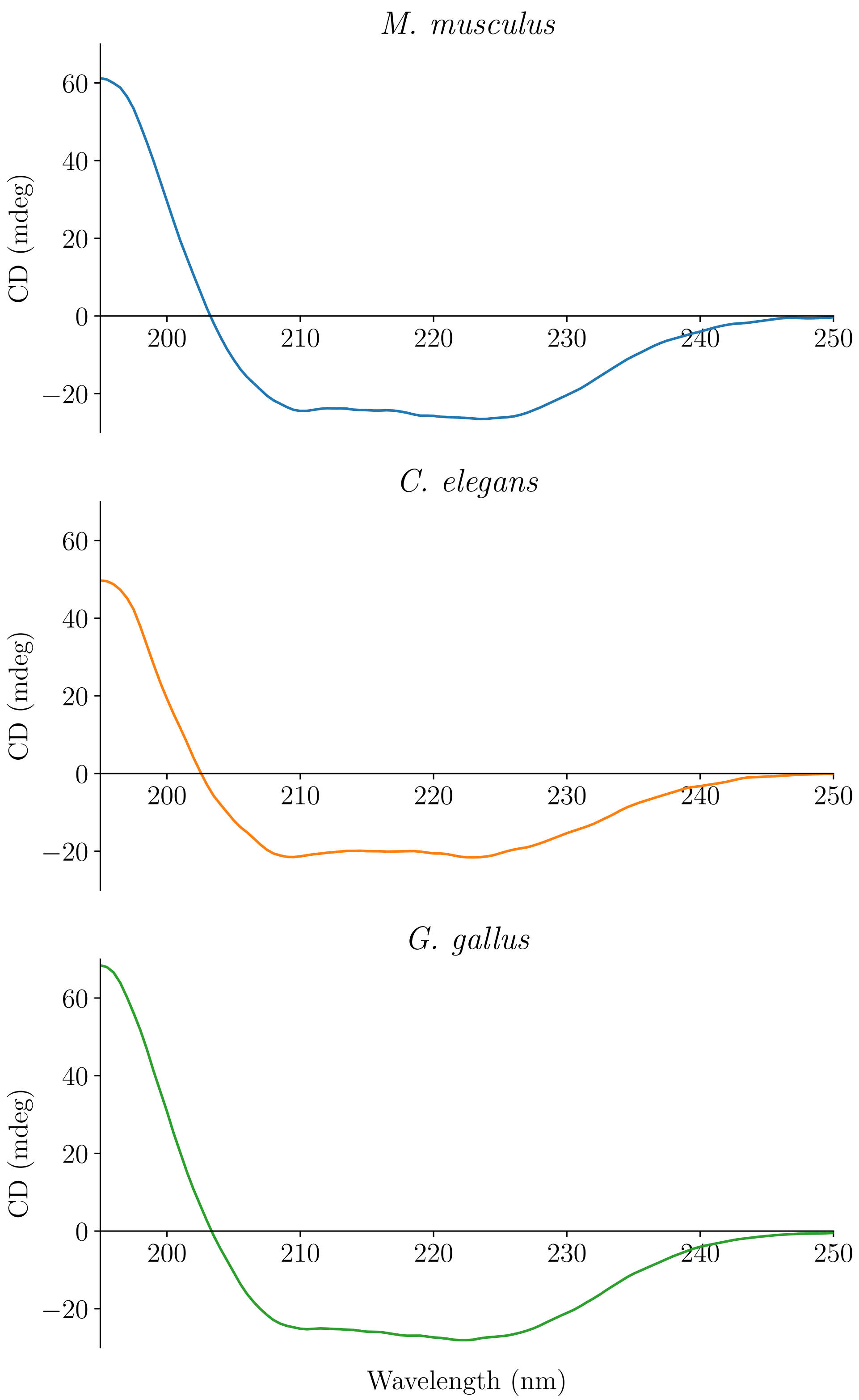


Figure 2.9 CD scan profiles of three KDELR homologues. *Mm* and *Gg* were measured in 20:1 DDM:CHS, while *Ce* was measured in 10:1. All measurements performed in 0.03% detergent with 10 mM potassium phosphate pH 5.0 and 50 mM sodium acetate.

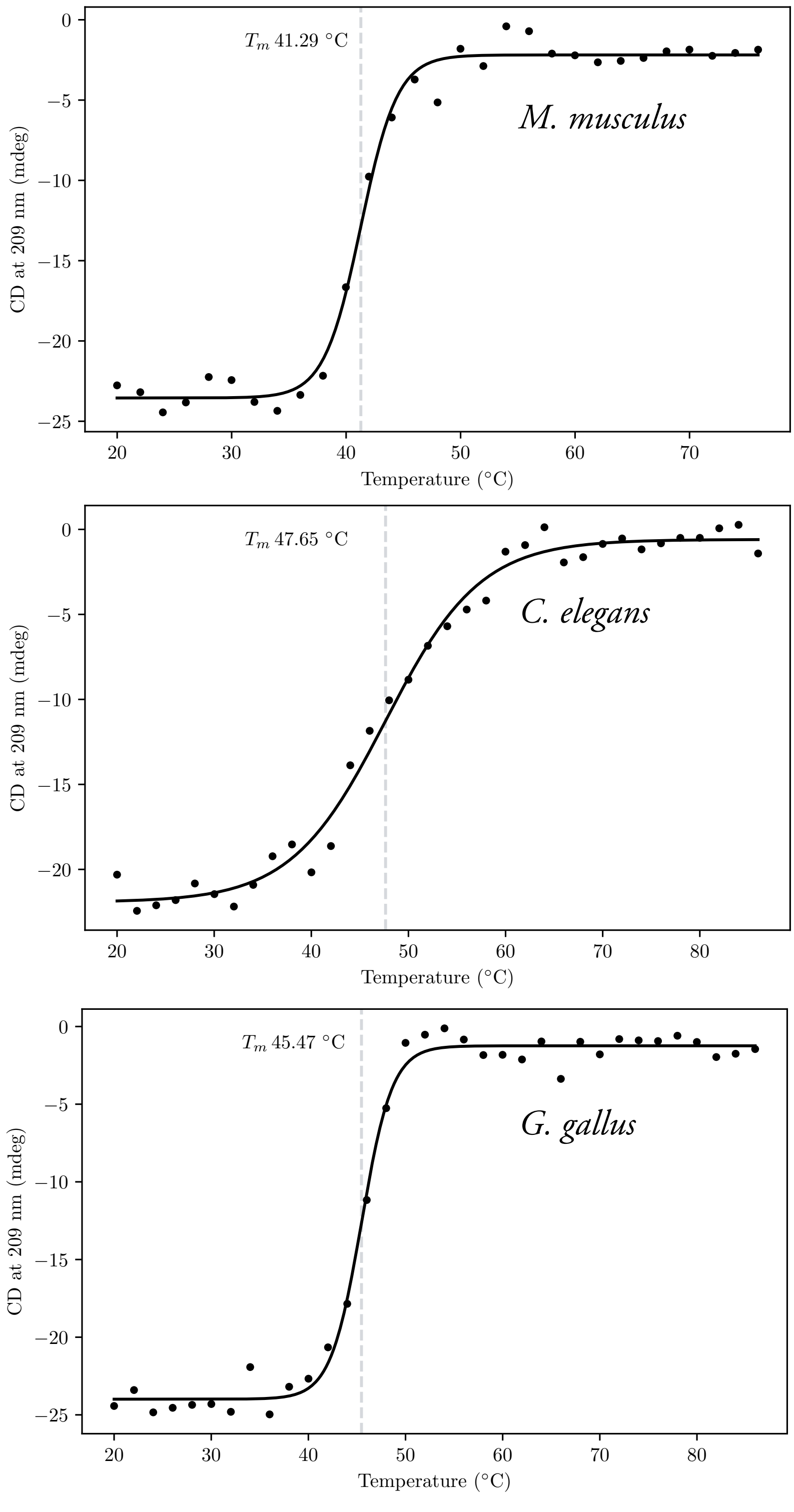


Figure 2.9 (Continued) CD melt profiles of three KDELR homologues. *Mm* and *Gg* were measured in 20:1 DDM:CHS, while *Ce* was measured in 10:1. All measurements performed in 0.03% detergent with 10 mM potassium phosphate pH 5.0 and 50 mM sodium acetate. The full profiles can be seen in figure B.1.

2.2 Peptide Binding

To determine if the purified KDELR showed specific ligand recognition properties, a peptide binding assay was developed. Initially CD melts and nanoDSF in the presence of a peptide (TAEKDEL) had indicated that binding was occurring as thermal stabilisation was observed (See appendix B, figure B.3). Thermal stabilisation in itself cannot however prove that binding is occurring and does not provide any kinetic information. A variety of other approaches to develop a binding assay were subsequently attempted ranging from simple pulldowns with proteins that had a C-terminal KDEL sequence, to microscale thermophoresis (MST), fluorescence polarisation (FP), Surface Plasmon Resonance (SPR), Bio-layer Interferometry (BLI), and finally a tritiated TAEKDEL peptide binding assay, which is described in the methods section 8.7.1. Figure 2.10 shows the result of a peptide binding assay performed with *Gg* KDELR against a TAEKDEL peptide, showing a low micromolar affinity at pH 5.4; confirming the binding activity of the purified protein, even in a detergent solubilised state.

2.3 Conclusion

After extensive screening *Gg* was determined to be the most promising candidate for crystallography. Due to its superior thermal stability (T_m >50 °C), presence of α-helical secondary structure, good yields (>1.5 mg from 6 L of culture), high purity (>90 %) and its ability to bind a KDEL peptide after purification (K_d 1.9 µM). The screening process was however not performed on all homologues at the same time, such that for a period of time, the most promising candidate was considered the *Mm* KDELR. This is the reason why it was used as a target for Sybody generation in chapter 3, Synthetic Nanobodies.

The extensive screening process performed here forms the basis of the successful structural studies performed subsequently, and lays the ground work for the design and execution of subsequent experiments. If protein crystallography

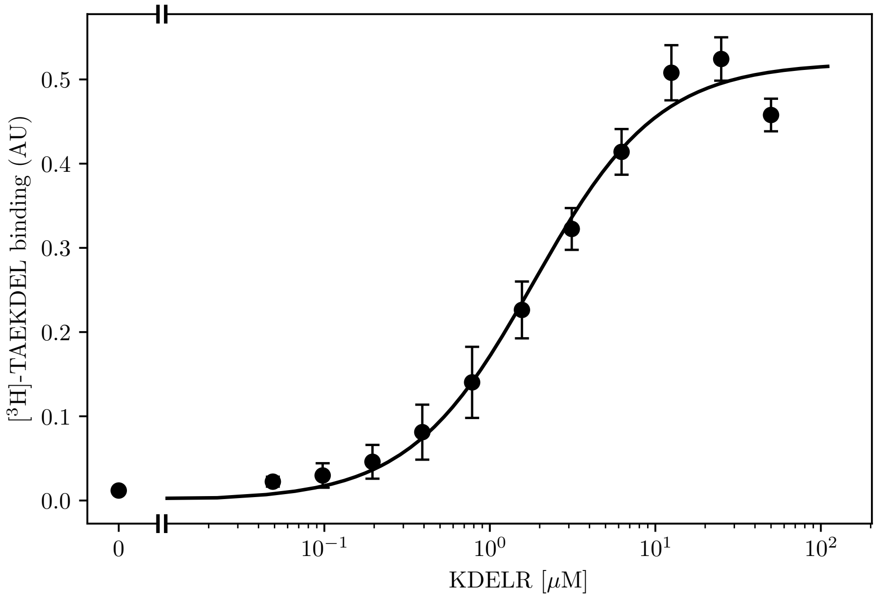


Figure 2.10 Binding curve of *Gg* KDELR with ^{3}H-TAEKDEL peptide at pH 5.4. Fit is to a three parameter sigmoid, fitting the maximum, slope, and inflection point. Apparent $K_d = 1.9 \pm 0.5$ µM, the hill slope is 1.19 ± 0.12 indicating non-cooperative binding. Experiment was repeated three times, each measured once. AU are arbitrary units normalised to the scintillation counts of the peptide added to the mixture, where 1 would represent a theoretical 100% binding of peptide added. Error bars are Standard Deviation (SD).

is like playing the lottery, then screening homologues is choosing the lottery with the best odds in which to play, and setting up many conditions is the purchase of thousands of tickets within that lottery. While playing in many lotteries with thousands of tickets increases the chances of success, other tricks are available to crystallographers to which lottery punters do not have access to. The generation of synthetic nanobodies to aid in crystallisation is described in the next chapter.

3 Synthetic Nanobodies

3.1 Nanobodies as Crystallisation Aids

Crystallisation of membrane proteins is a challenging endeavour. One of the methods that can be employed to facilitate such a difficult process is the development of stable binder molecules against the target protein. Crystallisation chaperones perform several tasks that are desirable from a crystallographers point of view: they bind and stabilise particular conformations of a protein, thus reducing the movement of the molecule, they provide crystal contacts, and they may potentially be used to estimate the phases of the data with molecular replacement. Single domain camelid immunoglobulins, known as nanobodies, are antibody fragments derived from a type of Immunoglobulin (Ig) with only one heavy chain and three Ig domains, that were first discovered in camels in 1993 (Hamers-Casterman et al., 1993). Nanobodies are one of several crystallisation chaperones derived from antibodies and they have several advantages over other small binders like antigen binding fragments (Fab), which are generally derived from papain cleavage of IgG or IgM antibodies (Mihaesco & Seligmann, 1968) or single chain variable fragments (scFv), which are a fusion of the variable domain of the heavy chain (VH) and variable domain of the light chain (VL) domains, see figure 3.1 (Bird et al., 1988). Full Igs are non-trivial to produce and are not suited for structural biology due to flexibility in their hinge regions and glycosylation (Ereño-Orbea et al., 2018; Desmyter et al., 2015). The use of Fabs over scFvs for the purposes of structural biology is generally due to their greater stability, especially at high concentrations (Koide, 2009).

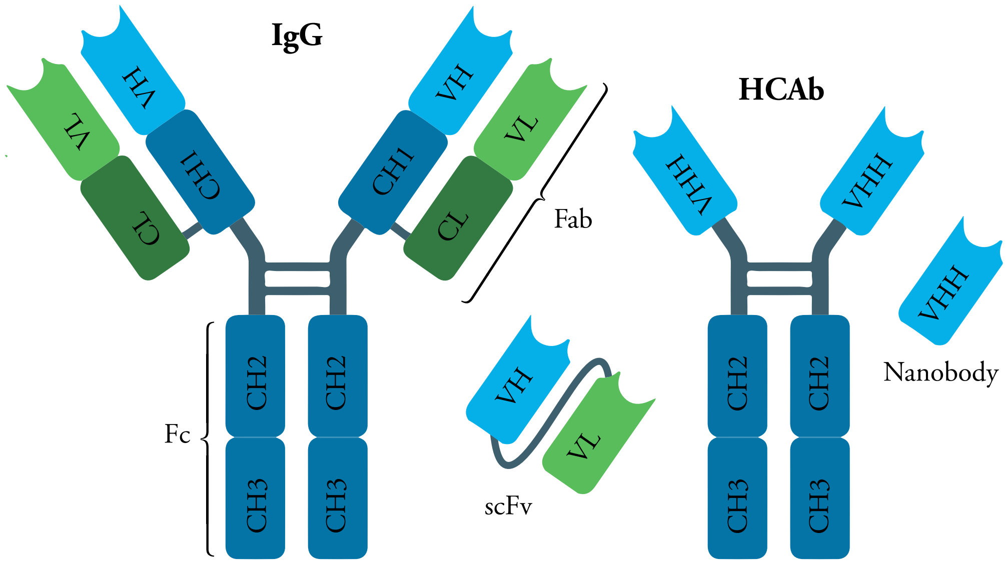


Figure 3.1 A cartoon depiction of a common IgG on the left, showing the heavy (blue) and light chains (green). Fab and scFv are shown in relation to IgG. A heavy chain antibody (HCAb) from camelids is shown on the right with its relation to the derived nanobody fragment.

Nanobodies present several features which make them the crystallisation chaperone of choice. Firstly, as the variable region is a single domain, all of the binding activity will be due to the single VHH. The simplicity of this system makes identification of the VHH relatively easy and allows Polymerase Chain Reaction (PCR) amplification from complimentary DNA (cDNA) in a single amplicon (Muyldermans, 2013). The single domain is thus also easily amenable to being cloned into vectors for phage display and can be expressed in bacterial cultures at high yields. An intriguing feature of nanobodies is their Complementarity Determining Region 3 (CDR3) domain, which is longer than classical CDR3s and thus shows a propensity to bind concave surfaces, sometimes inserting deeply into enzyme active sites (Desmyter et al., 2015). Nanobodies have also been shown to resist high temperatures (apparent T_m of 60–80 °C) and demonstrate high chemical stability (Dumoulin et al., 2002).

Despite the clear benefits nanobodies present, there are some challenges in their generation. Classically the generation of nanobodies is performed by first immunising a camelid, collecting its blood and preparing lymphocytes, extracting the messenger RNA (mRNA) and generating a VHH library for panning with phage display methods, though many variations of this process have been described (Muyldermans, 2013; Desmyter et al., 2015; Fridy et al., 2014). Some major restrictions are imposed on the process by immunising a camelid (Zimmermann et al., 2018a). Target proteins must be stable enough to remain folded upon injection both due to the adjuvants and the high camelid body temperature (38 °C). Targeting specific conformations is also very difficult, because ligands are likely to dissociate unless their binding is extremely tight. Finally, access to animals for this process is not always convenient (Zimmermann et al., 2018a).

In theory a purely *in vitro* selection process would allow for the selection of binders against specific ligand-bound conformations of the target protein in a defined, buffered environment. This would allow the generation of binders to specific states of the target protein in a precisely controlled manner. Unstable proteins with poor affinities to their ligands could be stabilised in their bound-state, by careful choice of buffer conditions and ligand concentrations, enabling the selection against a state that could not be replicated in *in vivo* selection. It is this level of control that is one of the chief benefits of *in vitro* selection.

3.2 Synthetic Nanobodies

To overcome the difficulties associated with immunisations, Zimmermann et al. developed an *in vitro* selection platform for libraries of synthetic nanobodies (Sybody). Some of the challenges that had to be overcome in order to develop a successful pipeline for *in vitro* selection include the potentially limited library sizes, poor framework design and selection bias (i.e. enrichment of binders unspecific to the target) due to the method of target display (Zimmermann et al., 2018a). To overcome these Zimmermann et al. used a combination of ribosome display and phage display techniques as well as a variation of target immobilisation methods, in conjunction with very stable sybody scaffolds. Their selection process can be performed within three weeks.

Library	PDB Template	CDR3 Length	No. randomised residues	Theoretical Diversity
Short	3K1K - GFP	6 aa	15	$8.3 \cdot 10^{17}$
Medium	3PoG - GPCR	12 aa	16	$4.3 \cdot 10^{19}$
Long	1ZVH - Lysozyme	16 aa	18	$2.8 \cdot 10^{22}$

Table 3.1 Selected properties of the three synthetic nanobody libraries developed by **Zimmermann et al.**.

Three libraries based on three binding modes observed in the PDB were designed and their scaffolds were shown to be extremely thermostable. The key differences between the three libraries are shown in table 3.1. The three libraries are termed concave, loop and convex or as they will be called here, short, medium and long and are seen in figure 3.2. The positions to be randomised were selectively enriched for certain amino acid combinations to further optimise the binders, prolines and cysteines were excluded from these regions entirely. This was done during the synthesis process by using specific tri-nucleotide mixes. One of the considerations was an observation that decreased surface hydrophobicity can be important to avoid the enrichment of sticky binders, which bind membrane proteins non-selectively though hydrophobic interactions (**Zimmermann et al., 2018a**).

As synthetic binder libraries cannot perform affinity maturation like the immune system of a camelid, it is important to have a large initial library diversity. The theoretical library diversity of the three designs certainly meets this criterion. However, commonly performed display methods such as yeast and phage are limited to 10^9–10^{10} members. Ribosome display methods can display up to 10^{12} library members and thus form the first selection stage (**Zimmermann et al., 2018a**).

The selection platform against membrane protein goes through the following general stages, selecting against each of the three libraries: ribosome display, two rounds of phage display, the second of which is performed with competition for binders from free (un-biotinylated) target protein to select for the tightest interactions, with the slowest off rates. Finally Enzyme-Linked Immunosorbent

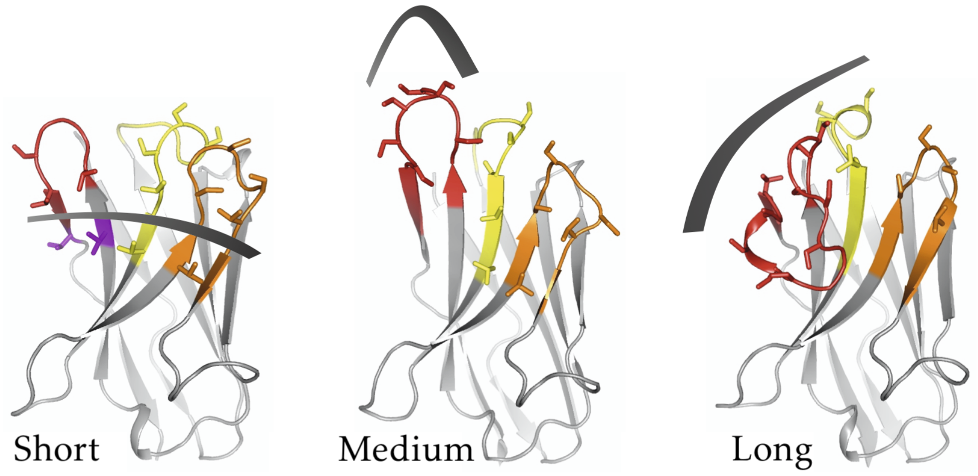


Figure 3.2 The three synthetic libraries based on thermostabilised nanobody frameworks. CDR1, CDR2 and CDR3 are coloured in yellow, orange and red, respectively. The different binding surface shapes are shown by the black line. Adapted from Zimmermann et al. (2018a).

Assay (ELISA) experiments are performed to confirm the presence of sybodies which are then test purified and screened with SPR. The use of different display systems (ribosome and phage) should remove background binder selection, furthermore the display processes are performed with alternating target immobilisation (streptavidin T1 and C1 Dynabeads, and neutravidin coated maxi-sorp plates) thus presenting different surface chemistries for every round, in order to avoid selection bias and background binders. In addition to this, solution panning was carried out, so display particles were incubated with target in solution followed by a subsequent target-pull down (Zimmermann et al., 2018a).

3.3 Selection against KDELR

Given the potential of sybodies, we set out to select binders against the *Mm* KDELR. Two conformations of the KDELR were to be selected for, the unbound, apo state at a neutral pH and a peptide-bound state at an acidic pH, as binding is pH-dependent and maximal at pH 5 (Wilson et al., 1993). With the goal of using the resulting sybodies to solve the crystal structures of the KDELR in those conformations and to subsequently test their effects on the KDELR retrieval system *in vivo*.

At this stage of the project the construct in use was the *Mm* KDELR. No binding assays had been successfully developed at the time - from the literature we deduced that the binding in liposomes was in the nanomolar range and while the human KDELR binds a KDEL C-terminal sequence, competition assays had show a small preference for a HDEL sequence in rat KDELR (**Wilson et al.**, 1993; **Scheel & Pelham**, 1996). Given the sequence similarity between the human, mouse and rat KDELR proteins the two conditions were decided as: 50 nM *Mm* KDELR at pH 7.5 for the apo state and 50 nM *Mm* KDELR with 200 nM of a TAEHDEL peptide at pH 5.0 for the peptide-bound state.

3.3.1 Sample Preparation

In order to immobilise the KDELR for panning, a covalently attached biotin is required. This was achieved by inserting an Avi-Tag into the KDELR expression vector in between the KDELR sequence and the C-terminal TEV protease cleavage site. An Avi-Tag is a 15 amino acid peptide sequence that is specifically recognised and biotinylated by the *E. coli* enzyme BirA. Avi-tagged KDELR was purified and subsequently incubated with BirA enzyme, which biotinylates a specific lysine residue in the Avi-Tag sequence via an ATP intermediate (biotinyl 5'-adenylate). After biotinylation and a subsequent purification SEC to remove BirA and any excess biotin, a streptavidin shift assay confirmed full biotinylation of the KDELR, seen in figure 3.3.

3.3.2 Ribosome Display

Commercially available *in vitro* translation mix was incubated with sybody library mRNA. The library mRNA are designed with 5' and 3' stem loops, stabilising the mRNA. After the sybody coding region a TolA based spacer is inserted before the 3' stem loop. As there is no stop codon the ribosomes stall, forming a complex of mRNA, ribosome and protein. This complex was stabilised by the addition of magnesium ions and reducing the temperature after the translation reaction had taken place. Ribosomal complexes were then incubated with *Mm* KDELR and subsequently these complexes were immobilised on streptavidin beads and washed. Eluted mRNA was reverse transcribed and amplified by PCR, a quantitative PCR (qPCR) step was also performed to ensure the reverse transcription process produced full length cDNA. The amplified cDNA was then cloned into the pDX_init vector for phage display.

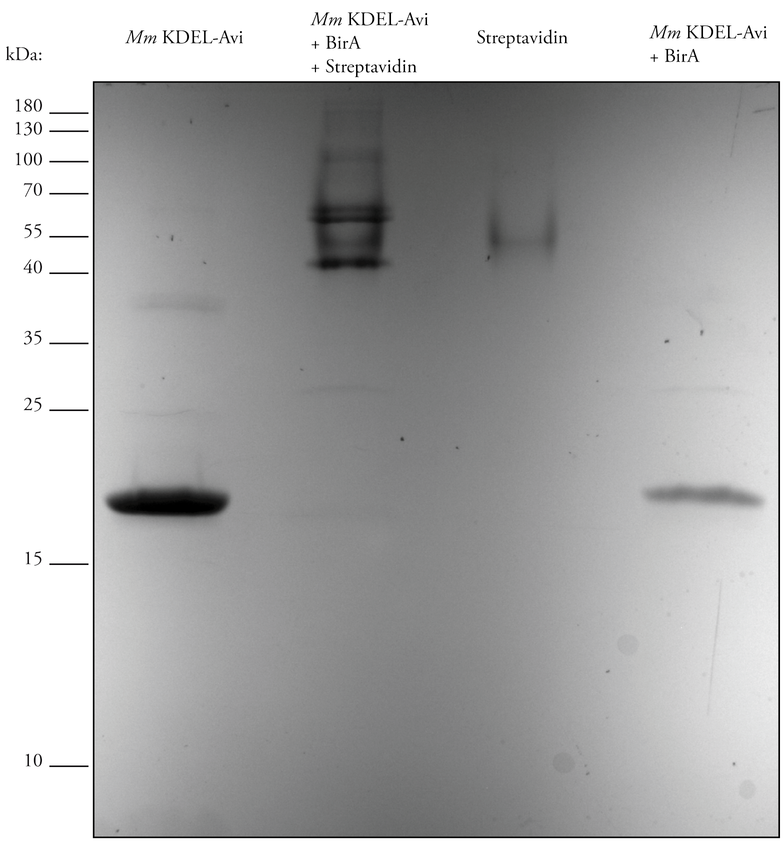


Figure 3.3 SDS-PAGE of *Mm* KDELR-Avi, in the first lane before incubation with BirA, in the second lane after incubation with BirA and post SEC, with streptavidin added in excess, in the third lane streptavidin alone, in the fourth lane after incubation with BirA and post SEC. In the second lane we observe a full shift of the *Mm* KDELR-Avi band, indicating complete biotinylation. Stained with coomassie brilliant blue.

3.3.3 Phage Display

Competent *E. coli* cells were transformed and phages produced overnight and precipitated from the culture supernatant. Phages were incubated with *Mm* KDELR and the complexes immobilised on a maxi-sorp plate coated in neutravidin. After washing, bound phages were eluted and used to infect a second batch of cells for the second round of phage display. The second round was performed with streptavidin-coated magnetic beads. The second round was performed in two different manners, once as above and once with competition for phage binding by un-biotinylated *Mm* KDELR to select for the tightest binders. In all rounds of phage display, each library was also bound in a control experiment to a biotinylated AcrB. This allowed the monitoring of the excess of polynucleotides eluted in a round of display, comparing the selection against the target and the control, which was not selected against. As can be seen in table 3.2 the enrichment for selection in low pH conditions, selecting for a peptide-bound state, was poor, so in this case it was decided to continue with the pool of phages from the second round that was performed without competition from un-biotinylated *Mm* KDELR.

		Enrichment		
	Libary	Round 1	Round 2	Round 2, Competition
pH 5.0	Small	0.4	0.7	1.2
	Medium	0.4	1.4	0.9
	Large	0.3	0.8	0.7
pH 7.5	Small	1.6	5.5	7.8
	Medium	1.2	4.0	4.9
	Large	0.7	2.6	3.6

Table 3.2 Enrichment of polynucleotides from phage display when comparing target selection to selection against an off-target control protein (AcrB), as monitored by qPCR.

3.3.4 ELISA

In order to validate sybody clones that were subcloned from phagemids of round two into an expression vector, ELISAs were performed. Sybodies were immobilised onto plates via a C-terminal myc-tag in the crude cell lysate. After washing, biotinylated *Mm* KDELR was added and finally streptavidin-HRP. The results of these ELISAs can be seen in figures 3.4 and 3.5. The background levels of the negative control in the ELISAs for the peptide-bound selection are much higher than those performed at a higher pH and there were only few hits that were carried forward to the next step. The signal intensity of the ELISA has been found to have no correlation with binding affinity in this context, so signal was compared only the background levels, with an intensity over 1.5x background indicating a hit.

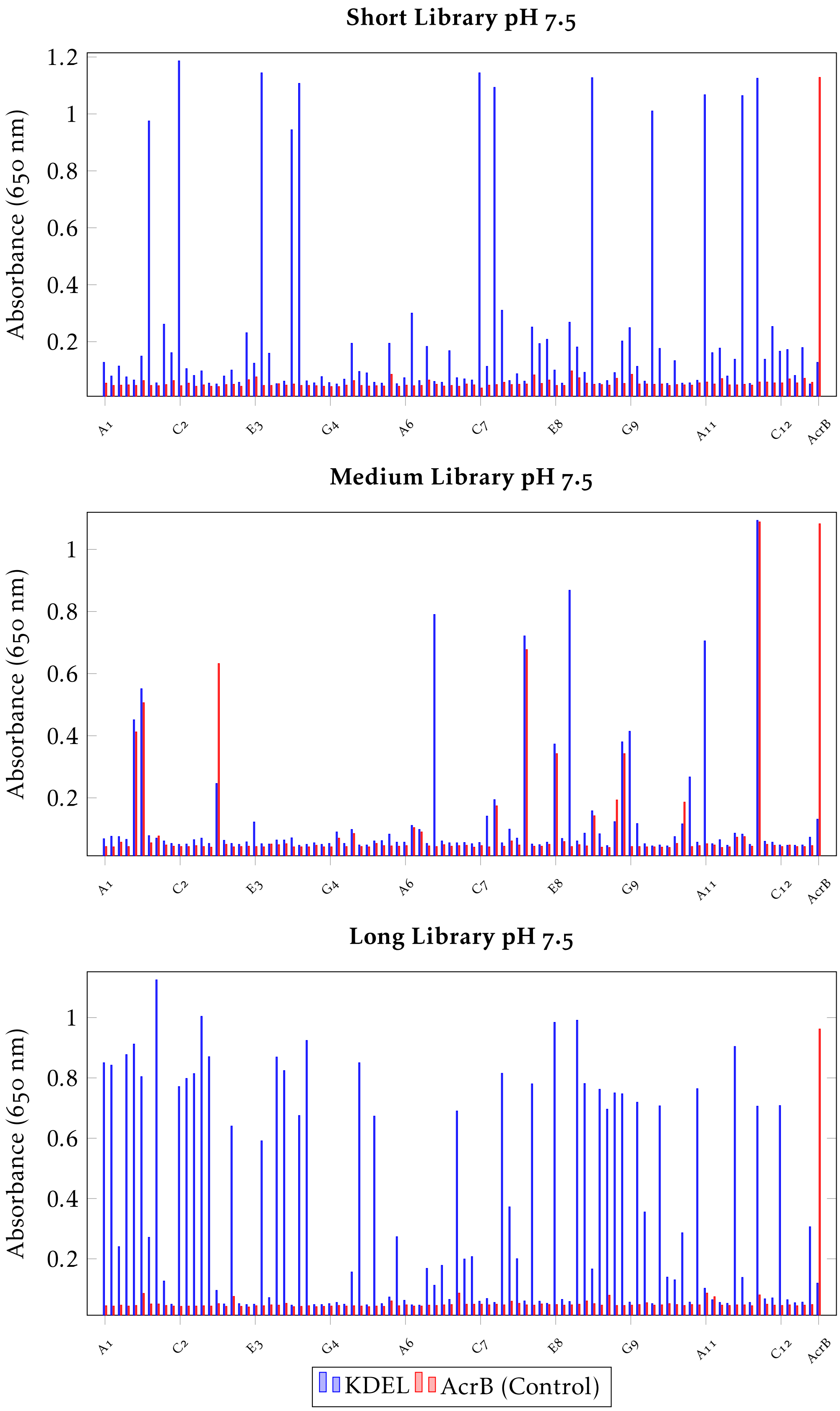


Figure 3.4 ELISA results of the screening of libraries at pH 7.5. The final value in each plot is the positive control against AcrB.

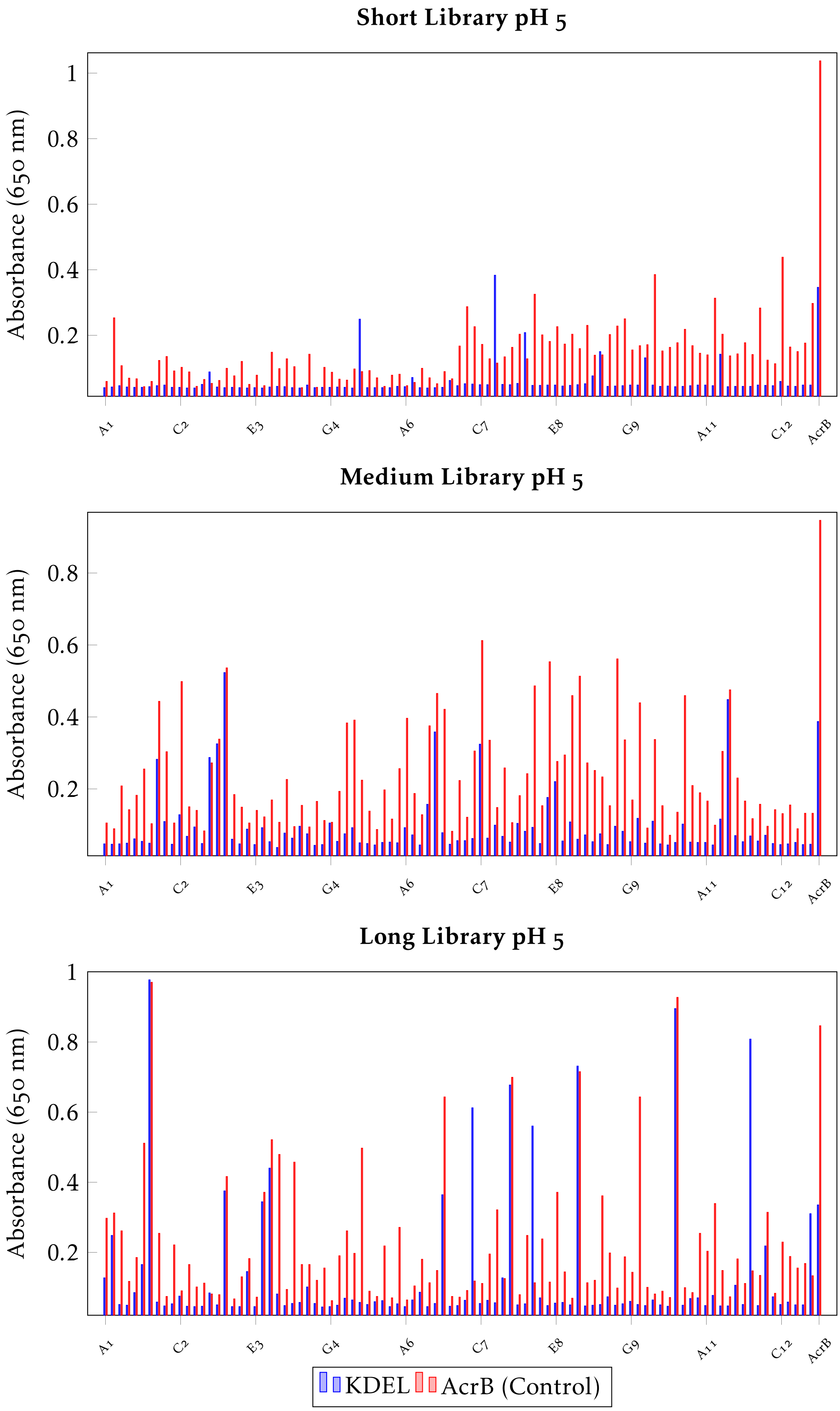


Figure 3.5 ELISA results of the screening of libraries at pH 5.0. The final value in each plot is the positive control against AcrB.

3.3.5 Purification & Screening

Binding sybodies identified from the ELISA screen were cultured on a small scale and sequenced, with duplicates being discarded. The remaining unique clones were expressed on a small scale and purified in two steps, Ni-affinity and SEC after osmotic lysis. The 16 unique, sybodies that could be purified are shown aligned in figure 3.6. Figure 3.7 shows the SEC profiles of all the sybody purifications attempted.

In order to determine if the sybodies were selective to a particular pH and if the presence of peptide at the low pH had an effect on the binding of the sybodies to the *Mm* KDELR, another ELISA experiment was performed with the purified sybodies (figure 3.8). As the sensitivity of an ELISA is limited, only large differences in affinity would be detected, none of these appear to be unequivocally detected. The addition of peptide at the low pH, to simulate the bound state, appears to have no effect on the ELISA results.

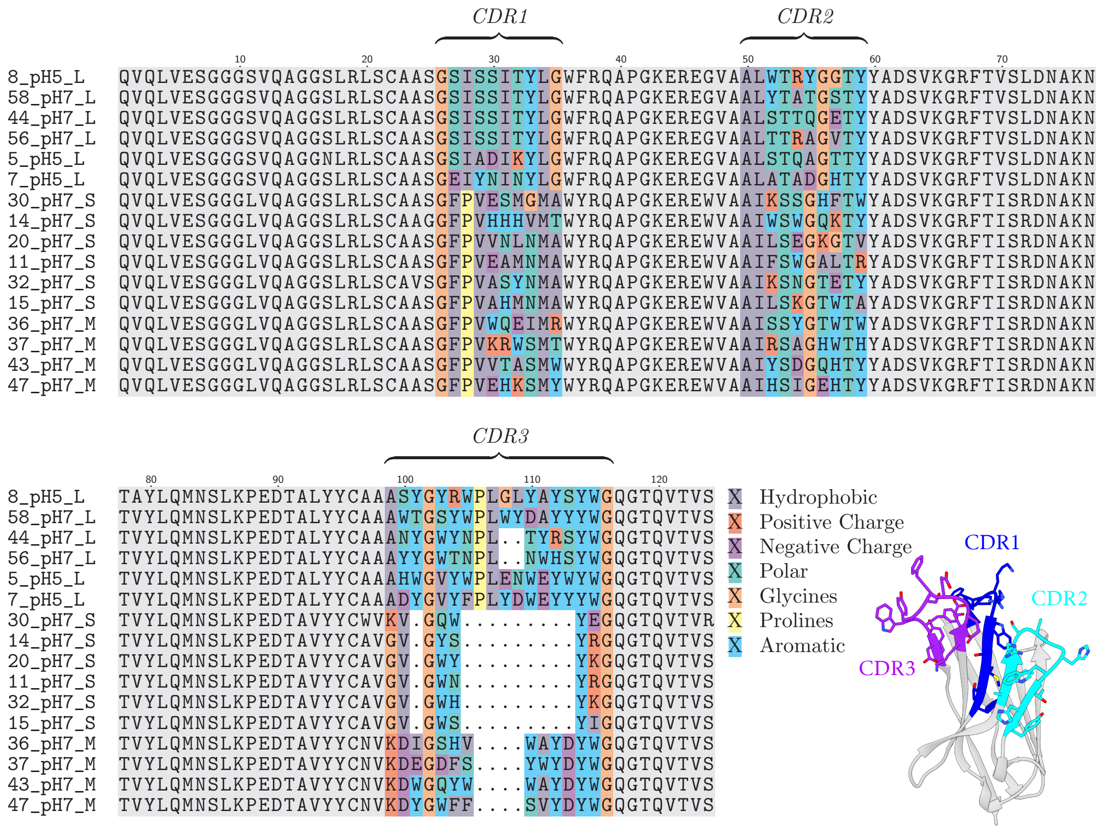


Figure 3.6 Alignment of the final 16 Sybodies that were successfully purified. In the bottom right a sybody is shown highlighting the CDR1, 2 and 3 regions on the structure.

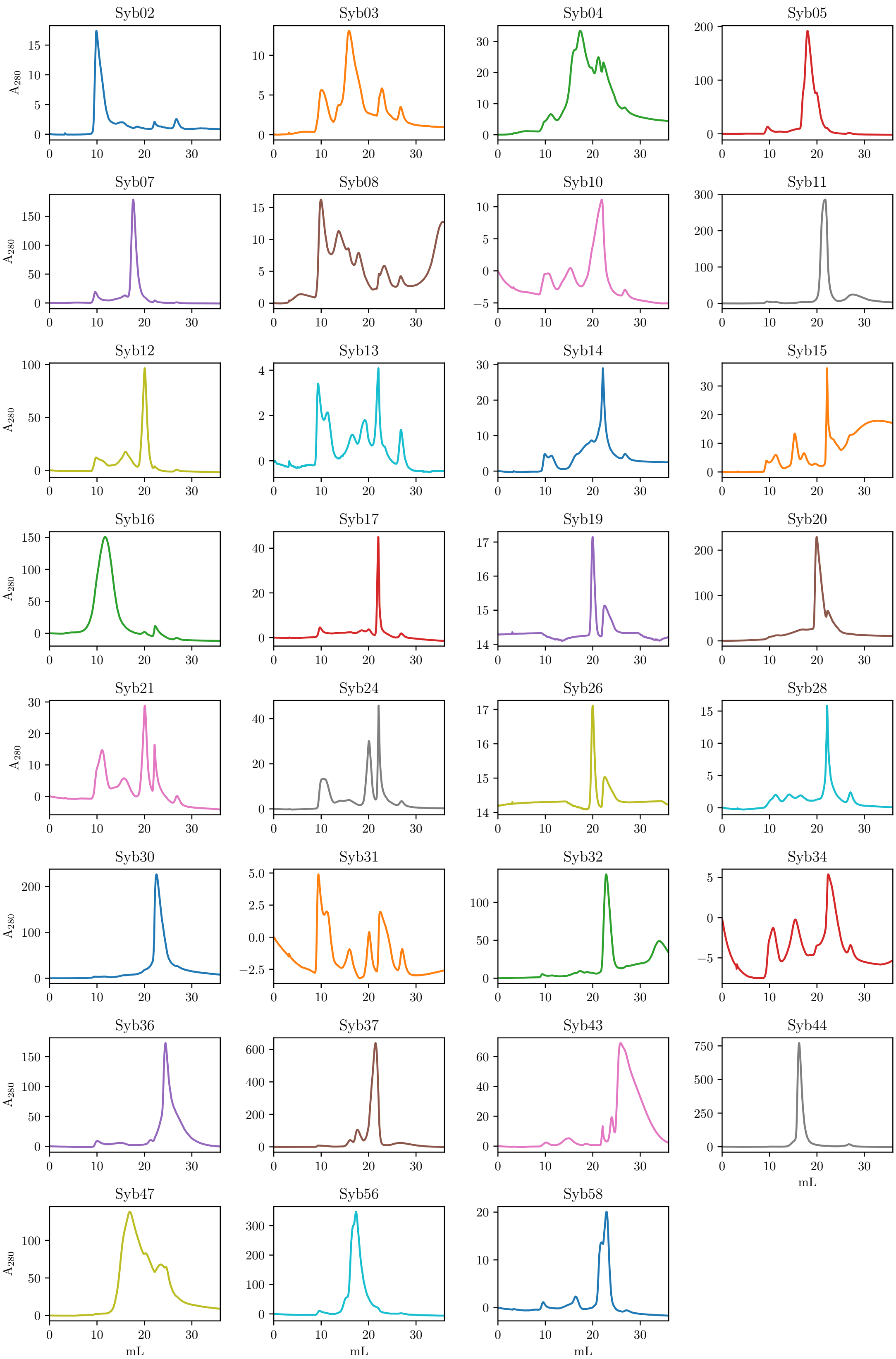


Figure 3.7 SEC traces after small scale purification of candidate sybodies.

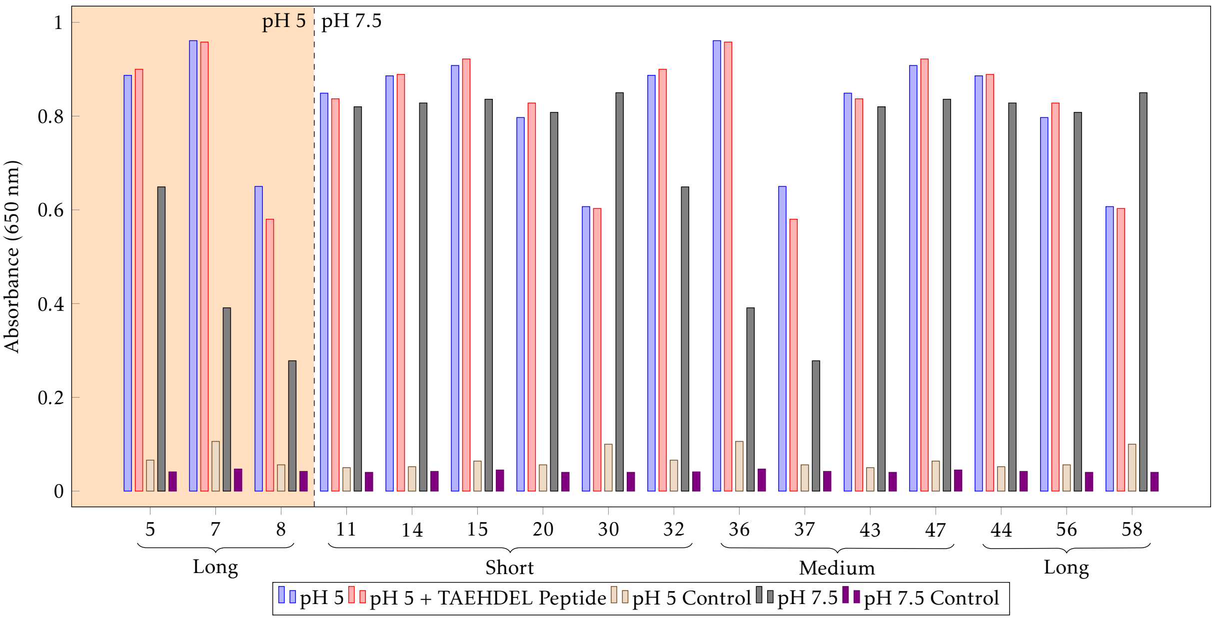


Figure 3.8 ELISA of selected sybodies after purification in two conditions, low pH with peptide and high pH without peptide.

Surface Plasmon Resonance

In order to further screen and dissect the binding properties of the sybodies, SPR experiments were performed to screen for binding in a more sensitive manner than afforded by an ELISA. The sybodies with a visible binding response are shown in figure 3.9. One sybody, Syb37 appears to have the highest affinity in this screen, with an estimated K_d of 15 nM.

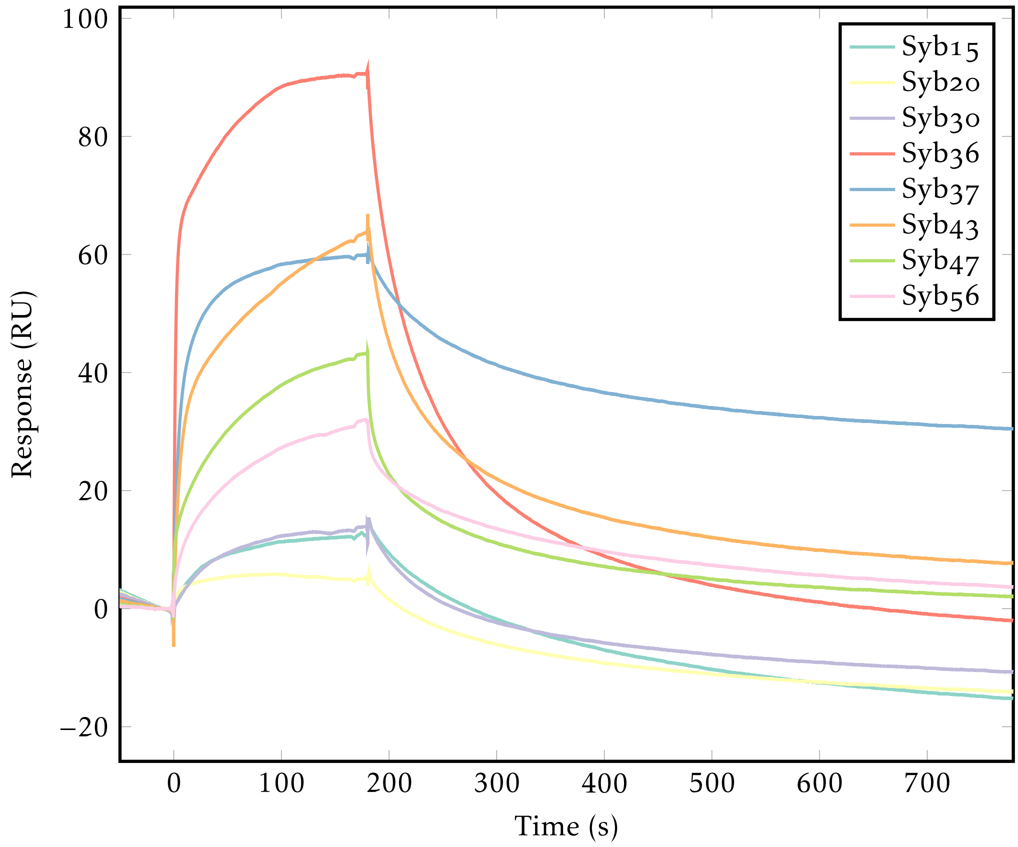


Figure 3.9 The SPR binding traces of the nine sybodies that showed a binding signal against *Mm* KDELR. All sybodies were screened at 1 µM in 20 mM Tris pH 7.5, 150 mM NaCl 0.03% DDM, 0.0015% CHS at 20 °C.

Size Exclusion Screening

Given the SPR results the next step was to determine if a stable sybody-KDELR complex could be formed and remain bound over the course of a SEC experiment. Complexes that succeed in this test are ideal candidates for crystallography as they appear to be very stable and can be purified as a complex with neither component in excess. The small weight of the sybodies mean that peak shifts are

pH	Family	Name	SPR	SEC-Shift	SEC-Coelution
5.0	Small	Syb5	Negligible		
		Syb7	Negligible		
		Syb8	Negligible	None Visible	
7.5	Small	Syb11	Negligible		
		Syb14	Negligible	None Visible	
		Syb15	Weak		
		Syb20	Negligible	None Visible	
		Syb30	Weak	None Visible	
		Syb32	Strong	None Visible	No
	Medium	Syb36	Intermediate	Shift	Yes
		Syb37	Strong	Shift	Yes
		Syb43	Strong	None Visible	No
		Syb47	Intermediate	None Visible	Yes
	Large	Syb44	Weak		
		Syb56	Intermediate	None Visible	
		Syb58	Strong	None Visible	Yes

Table 3.3 Screening results of sybody library against *Mm* KDELR. The sixteen sybodies in the table are those that could be expressed and showed binding in the ELISA experiments. SPR classification was based on the response units remaining compared to base line after 100 s of dissociation and estimated K_d from a 1:1 fit. Strong indicates an estimated K_d less than 100 nM with signal after 100 s of dissociation, intermediate indicates an estimated K_d less than 200 nM with signal after 100 s of dissociation, weak indicates an estimated K_d worse than 400 nM, and negligible indicates very poor if any K_d estimate with no signal after 100 s of dissociation. Blank spaces indicate the experiment was not performed.

difficult to discern on the SEC, so the peak of interest was run on an SDS-PAGE.

The results of this screening process are summarised in table 3.3.

3.3.6 Screening against Gg KDELR

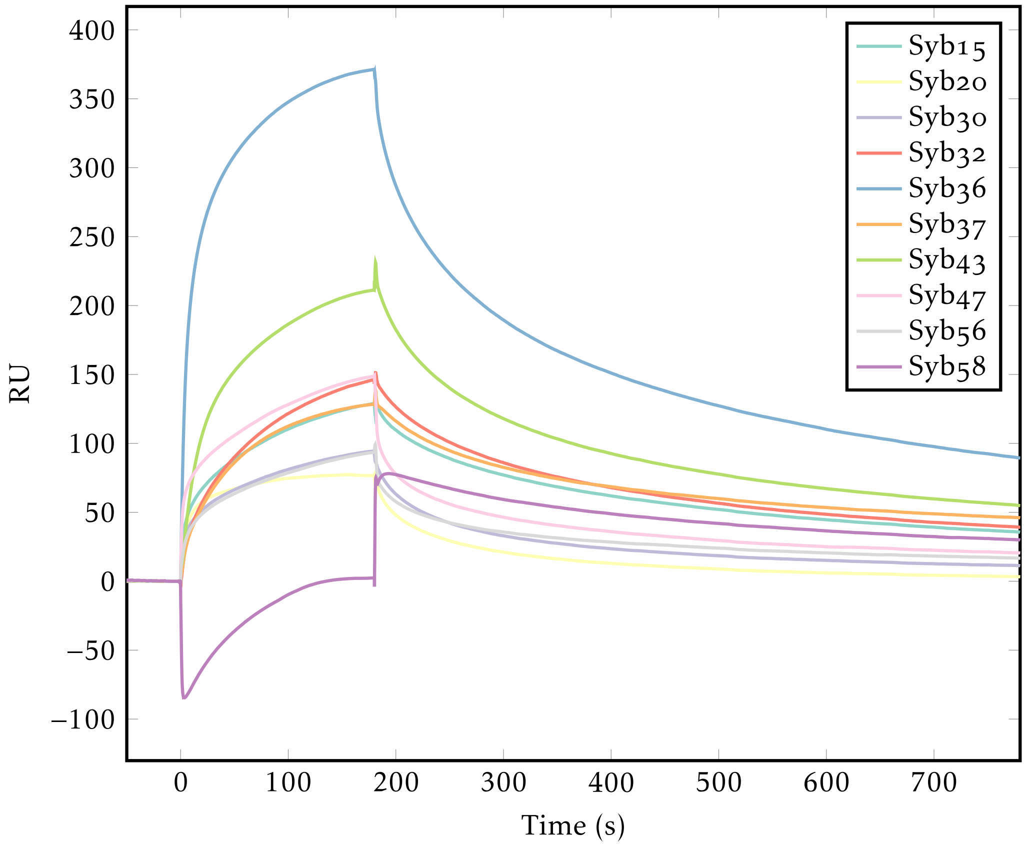


Figure 3.10 SPR of sybodies against *Gg* KDELR, the spike seen at Syb58 is due to a buffer mismatch. All sybodies were screened at 1 µM in 20 mM Tris pH 7.5, 150 mM NaCl 0.03% DDM, 0.0015% CHS at 20 °C.

At the same time as the sybody screening process was being performed against *Mm* KDELR, several other new homologues of the KDELR were being tested for stability. One of these, the *Gg* KDELR appeared promising. Given that the sequence identity between the *Mm* and *Gg* KDELR is 83.8%, with 93.3% similarity (figure 3.11), the sybodies shown to have some binding activity against *Mm* KDELR were screened against *Gg* KDELR. As seen in figure 3.10 and summarised in table 3.4, several of the sybodies selected against the *Mm* KDELR were able to bind *Gg* KDELR. The four tightest binders to *Mm* KDELR were Syb32, 37, 43 and 58 while the four tightest *Gg* KDELR binders were Syb36, 37, 43 and 58. The behaviour of the sybodies binding *Mm* and *Gg* is remarkably similar and indicates that the two proteins are bound in a similar fashion.

Family	Name	SPR	SEC-Coelution
Small	Syb15	Strong	
	Syb20	Weak	
	Syb30	Weak	
	Syb32	Strong	
Medium	Syb36	Strong	Yes
	Syb37	Strong	Yes
	Syb43	Strong	
	Syb47	Intermediate	
Large	Syb56	Weak	No
	Syb58	Intermediate	No

Table 3.4 Screening results of sybody library against *Gg* KDELR. SPR classification was based on the response units remaining compared to base line after 100 s of dissociation and the estimated K_d from a 1:1 model. Strong indicates an estimated K_d under 100 nM with signal present after 100 s of dissociation, intermediate a K_d worse than 100 nM with signal present after 100 s of dissociation, weak indicates that there was almost no signal remaining after 100 s of dissociation. Blank spaces indicate the experiment was not performed.

```
                     10        20        30        40        50        60
H. sapiens   MNLFRFLGDLSHLLAIILLLLKIWKSRSCAGISGKSQVLFAVVFTARYLDLFTNYISLYN
M. musculus  ............................................................
G. gallus    ..i..lt......a...i...................l...l...t.......sf.....

                     70        80        90       100       110       120
H. sapiens   TCMKVVYIACSFTTVWLIYSKFKATYDGNHDTFRVEFLVVPTAILAFLVNHDFTPLEILW
M. musculus  ................m...........................................
G. gallus    .s..li.....ya..y...m..................i..vgg.s.......s......

                    130       140       150       160       170       180
H. sapiens   TFSIYLESVAILPQLFMVSKTGEAETITSHYLFALGVYRTLYLFNWIWRYHFEGFFDLIA
M. musculus  ............................................................
G. gallus    .................i..........t....f..l..a...v......y.........

                    190       200       210
H. sapiens   IVAGLVQTVLYCDFFYLYITKVLKGKKLSLPA
M. musculus  ................................
G. gallus    v...v.............v.............
```

Figure 3.11 Sequence alignment of *Hs* KDELR1, *Mm* KDELR1 and *Gg* KDELR2.

3.4 Crystallography

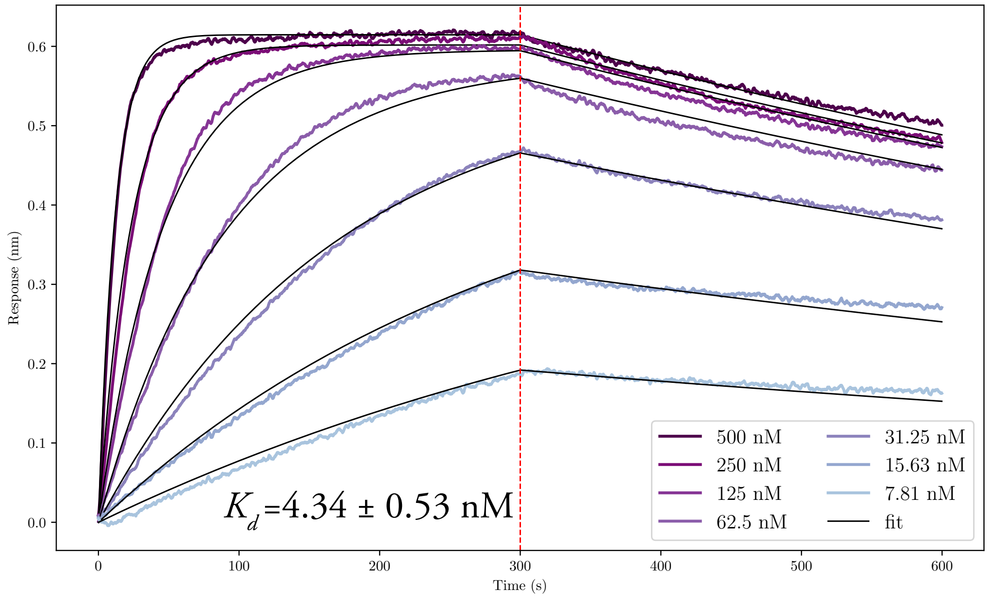


Figure 3.12 BLI measurement of Syb37 binding to *Gg* KDELR, the K_d as determined by kinetic analysis with a 1:1 binding model, is 4.34 ± 0.53 nM with an $R^2 = 0.998$, $n = 3$.

As the number of sybodies remaining after the screening process was limited, several were purified on a large scale (see figure 3.13 & 3.14) and set up in crystallisation trials with *Mm* and *Gg* KDELR in LCP. Those that co-eluted in a 1:1 stoichiometry from SEC (as judged by band density on SDS-PAGE) were set up from concentrated SEC fractions of the complex run, SDS-PAGE of *Mm* and *Gg* KDELR with Syb37 are seen in figure 4.11 and 4.12. Others were set up as co-crystallisation experiments in a 1:1 ratio. Only Syb37 formed a complex that was fully stable after SEC and only Syb37 formed crystals, both for *Mm* and *Gg* KDELR. Syb37 binds *Gg* KDELR with a K_d of 4.3 ± 0.5 nM, as determined by BLI (figure 3.12). The electron density of the *Mm* KDELR Syb37 crystals was very poor and this data was abandoned in favour of the high resolution maps obtained with *Gg* KDELR. An analysis of the interaction between Syb37 and *Gg* KDELR can be seen in section 5.2.

Family	Name	Purification	*Mm* Crystal	*Gg* Crystal
Small	Syb15	Impure		
	Syb32	Pure	No	No
Medium	Syb36	Pure	No	No
	Syb37	Pure	Yes	Yes
	Syb43	Precipitates		
	Syb47	Impure		
Large	Syb56	Pure	No	No
	Syb58	Pure	No	No

Table 3.5 List of all the sybodies purified on a large scale and the final state of each.

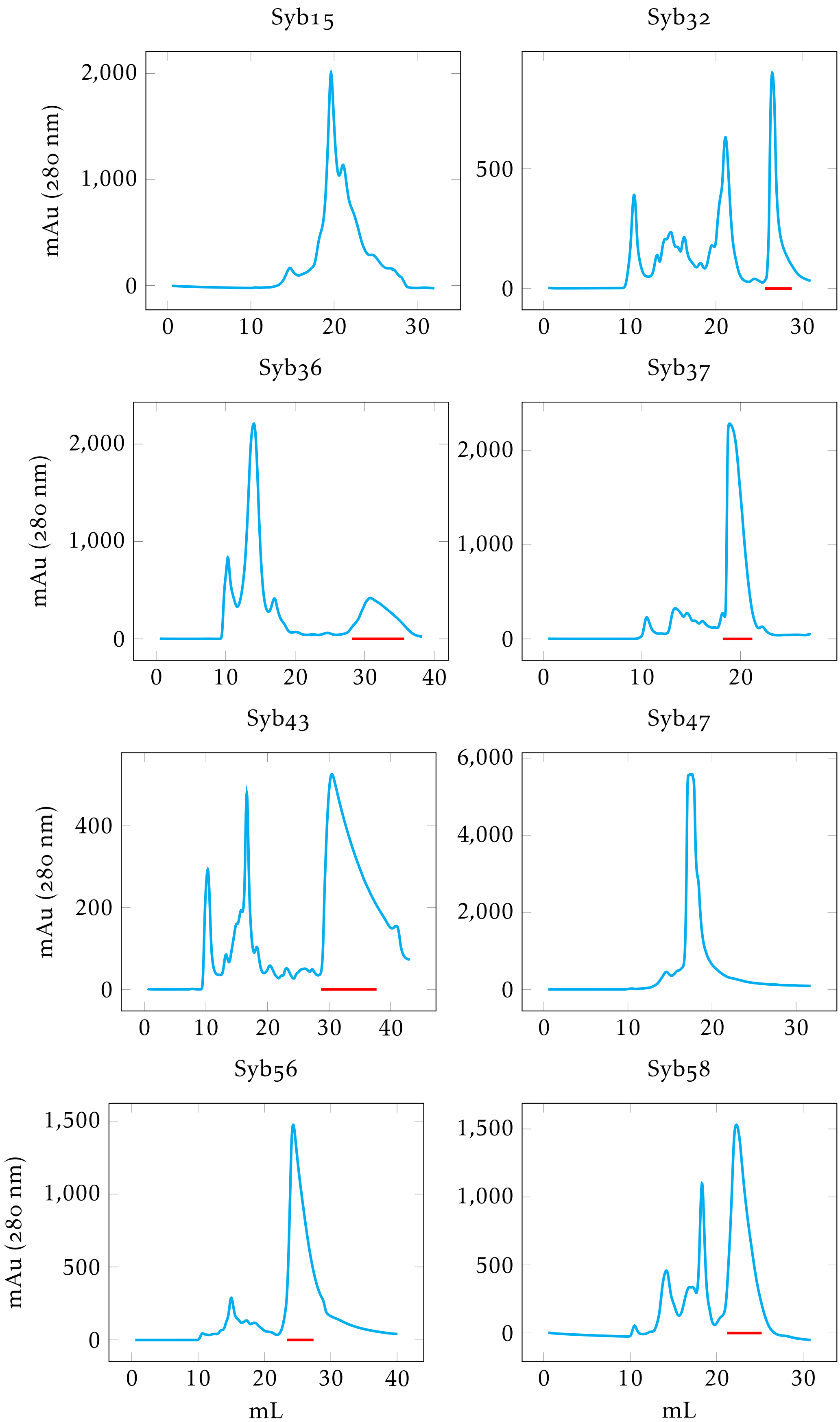


Figure 3.13 SEC traces of the purification of all Sybodies used for crystallography, on a Sepax SRT-C SEC 300 column. The red lines indicate the regions in the SDS-PAGE in figure 3.14. Column was not calibrated with molecular weight standards.

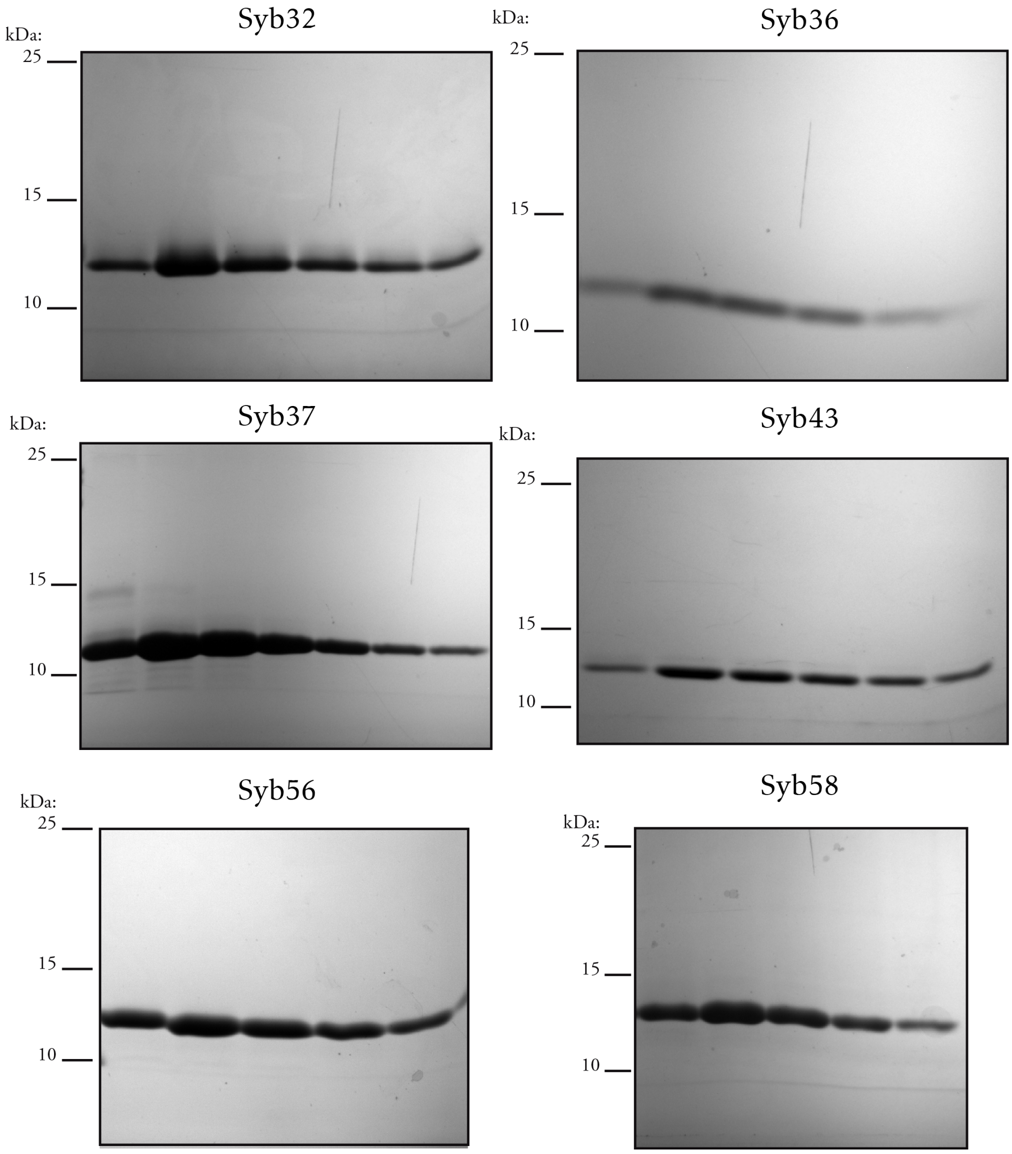


Figure 3.14 SDS-PAGE of the sybodies that could be purified successfully. The red lines in figure 3.13 indicate the regions in the SDS-PAGE gel. Stained with coomassie brilliant blue.

3.5 Conclusion

The screening platform was successfully employed to generate synthetic nanobodies which display characteristics that should facilitate crystallisation of the KDELR. Given the small (25 kDa) size of the KDELR, the success of this endeavour was not guaranteed and demonstrates the power of the approach developed by Zimmermann et al.. As figure 3.6 demonstrates, selected sybodies appear to be enriched in aromatic residues, with no obvious similarity to the K-D-E-L peptide sequence in any CDR. As the KDEL signal is C-terminal, this is not entirely unexpected, as a CDR loop would have to enter and exit the peptide binding site. Of course at the selection stage the epitope of the sybodies on the KDELR was not known, binding could be occurring to either the luminal or cytoplasmic faces of the KDELR.

In the search for sybodies the most time-consuming part concerns the screening of the pipeline output. In this case, only 16 sybodies had to be characterised. Rapid kinetic analysis of binders to membrane proteins is not trivial. One SPR run of 16 nanobodies at one concentration can take many hours. Given the slow throughput and then breakdown of the local SPR machine, a switch was made to BLI to determine the K_d of Syb37 binding. BLI implements a dipstick like method, where the ligand is immobilised on needles which are then submerged into wells containing analyte on a 96 or 384 well plate, instead of flowing sample across a sensor like SPR. This allows for many measurements in parallel, which also enables much faster condition optimisation. There is some potential for use of BLI directly from cell lysate, or even the result of the small scale purification. While measurements taken with different sybody concentrations cannot be compared by K_d directly, K_{off} is concentration independent and should offer some kinetic differentiation which is currently not possible with the ELISA. Once the final screening processes can be optimised and parallelised it is conceivable that a well characterised sybody library could be generated within four to five weeks.

Crystallography of the KDELR

4

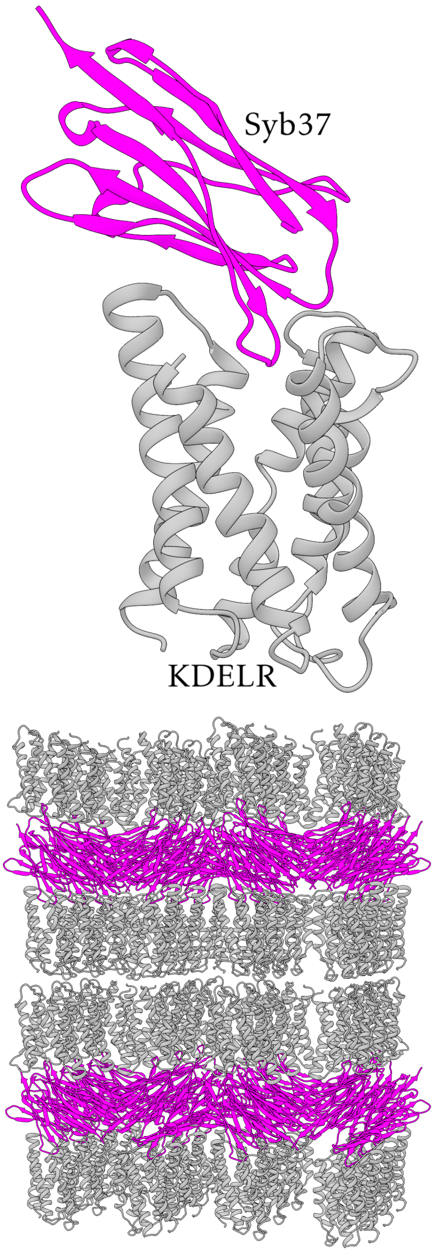
Crystal structures of the KDELR were determined at a high resolution in several states. The different structures will be briefly introduced followed by a description of the crystallographic processes used to obtain the data and build the models.

While all of the crystals grew at a similar point in time, it was the determination of the structure in complex with Syb37 that allowed the solution of the others. Homology models of the KDELR were insufficient for molecular replacement (MR). A model of a previously solved sybody could however be found in the KDELR-Syb37 dataset. From this it was possible to build into the remaining density and solve the complete structure of the KDELR. It was this structure that was then used as a search model for MR for the next datasets without Syb37. Initially while only *Gg* KDELR formed crystals in the apo or peptide-bound state, *Mm* KDELR-Syb37 complex formed poorly diffracting crystals. It was the poor quality of these which lead to the discovery that *Gg* KDELR-Syb37 also formed crystals, which diffracted to a very high resolution. Given their role in enabling the solution of the subsequent structures, we shall begin with the KDELR-Syb37 structures.

4.1 KDELR-Sybody37 Complex

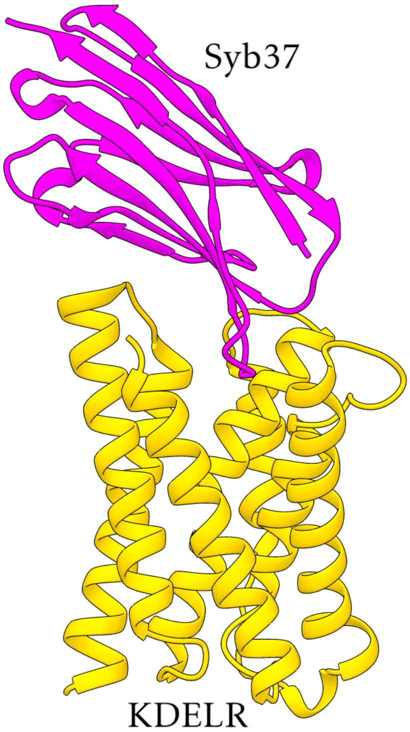
4.1.1 *Mm* KDELR-Syb37

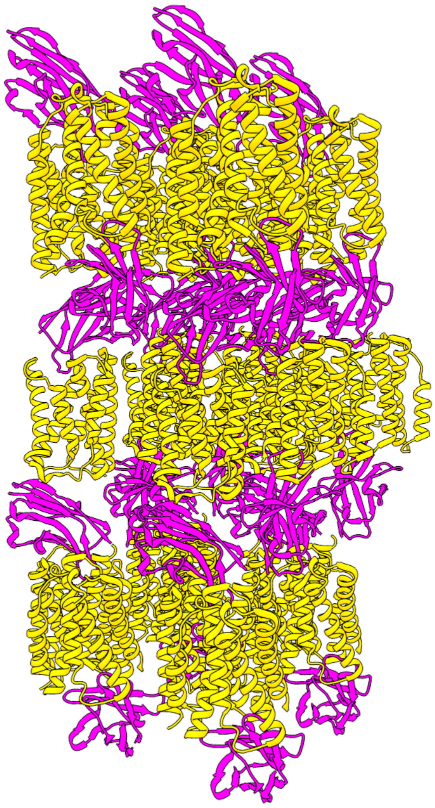
Before it was known that Syb37 could bind *Gg* KDELR, *Mm* KDELR-Syb37 crystals were grown and diffraction data was collected. *Mm* KDELR-Syb37 crystals grew in LCP under the following conditions: 20 °C, 30% (v/v) PEG



Mm KDELR-Syb37	
Data Collection	
Space Group	P 2_1 2_1 2
Cell Dimensions	
a, *b*, *c* (Å)	39.53, 94.41, 120.70
α, β, γ(°)	90, 90, 90
Wavelength (Å)	0.969
Resolution (Å)	47.20–3.09
	(3.214 – 3.103)
R_{pim}	0.055 (0.536)
R_{merge}	0.122 (1.247)
I/σI	7.0 (1.6)
Completeness (%)	99.32 (95.23)
Multiplicity	6.2 (6.3)
CC1/2	0.865 (0.536)
Wilson B-factor (Å²)	76.81
$\boldsymbol{B}_{cart}$ Eigenvalues (Å^2)	84.2, 57.7, 18.8

Figure 4.1 The Crystal structure of the *Mm* KDELR-Syb37 complex, crystal packing is shown below the asymmetric unit. KDELR is shown in grey and Syb37 is seen in pink. On the right the data collection statistics for this structure can be seen.





Gg xKDELR-Syb37 (6I6J)	
Data Collection	
Space Group	P 2_1 2_1 2_1
Cell Dimensions	
a, *b*, *c* (Å)	50.40, 52.98, 133.06
α, β, γ(°)	90, 90, 90
Wavelength (Å)	0.969
Resolution (Å)	47.13–2.23
	(2.29–2.23)
R_{pim}	0.055 (0.536)
R_{merge}	0.122 (1.247)
I/σI	7.0 (1.6)
Completeness (%)	99.5 (95.9)
Multiplicity	6.2 (6.3)
CC1/2	0.865 (0.536)
Wilson B-factor (Å²)	40.9
Refinement	
No. reflections	111218
Unique reflections	18108
R_{work}/R_{free}	0.21/0.25
No. atoms	
Protein	2638
(B_{res}), protein atoms (Å²)	49.1
(B_{res}), waters (Å²)	48.4
(B_{res}), lipid atoms (Å²)	67.9
Ramachandran favoured (%)	97.5
Ramachandran outliers (%)	0.0
R.M.S deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.95

Figure 4.2 The Crystal structure of the *Gg* KDELR-Syb37 complex, crystal packing is shown below the asymmetric unit. KDELR is shown in yellow and Syb37 is seen in pink. On the right the data collection and refinement statistics for this structure can be seen.

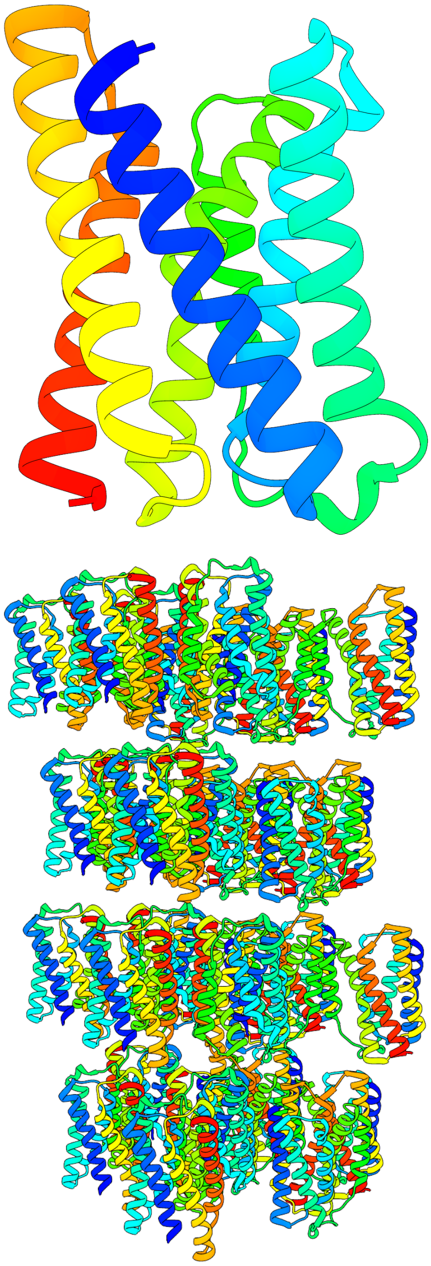
500 DME, 100 mM Tris pH 9.0, 100 mM sodium nitrate. The data from these crystals was of fairly low resolution (3.09 Å) with severe anisotropy (figure 4.1), the structure was thus not refined but rather used as a bootstrap MR for the *Gg* KDELR-Syb37 structure, which at this stage had not yet been set up.

4.1.2 *Gg* KDELR-Syb37

Like all the crystals structures presented here, *Gg* KDELR-Syb37 crystallised in LCP, under the following conditions: 20 °C, 30% (v/v) PEG 400, 100 mM Tris pH 9.0. As Syb37 was selected at pH 7.5 and the crystal formed at pH 9.0, one may postulate that this structure represents the state of the KDELR in the apo state, as found in the ER. This was later shown to be the case by the apo and peptide-bound structures. From the crystal packing the KDELR-Syb37 complex appears to be a monomer, with crystal contacts being made between Syb37-Syb37, KDELR-KDELR and KDELR-Syb37. CDR3 of Syb37 enters a pocket in the luminal side of KDELR, the only pocket in this structure. Data collection and refinement statistics can be seen in figure 4.2.

4.2 KDELR Apo Structure

The apo structure of the KDELR crystallised in LCP at 4 °C under the following conditions: 30% (v/v) PEG 500 DME, 100 mM Tris pH 9.0 and 100 mM magnesium sulphate. At this high pH and with no peptide present, this structure will most likely present the state of the KDELR in the ER. The structure appears incredibly similar to the KDELR-Syb37 structure, with an RMSD of 0.622 Å over 198 Cα atoms. As in the structure above, there is one potential binding pocket on the luminal side of the KDELR. Crystal packing once again indicates a monomeric state. Data collection and refinement statistics can be seen in figure 4.3.



KDELR (6I6B)	
Data Collection	
Space Group	$C\ 2\ 2\ 2_1$
Cell Dimensions	
a, b, c (Å)	47.90, 103.75, 101.59
α, β, γ(°)	90, 90, 90
Wavelength (Å)	0.969
Resolution (Å)	51.88–2.59
	(2.66–2.59)
R_{pim}	0.158 (0.348)
R_{merge}	0.359 (1.069)
$I/\sigma I$	4.1 (1.9)
Completeness (%)	99.9 (98.0)
Multiplicity	6.2 (6.3)
CC1/2	0.900 (0.519)
Wilson B-factor (Å²)	23.7
Refinement	
No. reflections	50410
Unique reflections	8183
R_{work}/R_{free}	0.24/0.27
No. atoms	
Protein	1680
(B_{res}), protein atoms (Å²)	30.8
(B_{res}), waters (Å²)	27.7
(B_{res}), lipid atoms (Å²)	51.0
Ramachandran favoured (%)	97.0
Ramachandran outliers (%)	0.0
R.M.S deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.96

Figure 4.3 The Crystal structure of apo KDELR, crystal packing is shown below the asymmetric unit. On the right the data collection and refinement statistics for this structure can be seen.

4.3 KDELR Peptide-Bound Structures

4.3.1 TAEKDEL Bound

The TAEKDEL peptide-bound structure of KDELR crystallised in LCP under the following conditions: at 20 °C 30% (v/v) PEG 600, 100 mM MES pH 6.0 and 100 mM sodium nitrate. Presence of the clearly resolved KDEL peptide in the binding cavity observed in the previous structures as well as the low pH, indicate that this structure represents the KDELR in its cargo-bound, Golgi state. Compared to the apo structure we see an RMSD of 0.760 Å over 165 pruned atom pairs but 2.058 Å across all 203 Cα pairs. Crystal packing again indicating a monomeric state, data collection and refinement statistics can be seen in figure 4.4. A larger OMIT map of the peptide is shown in figure 4.7

4.3.2 TAEHDEL Bound

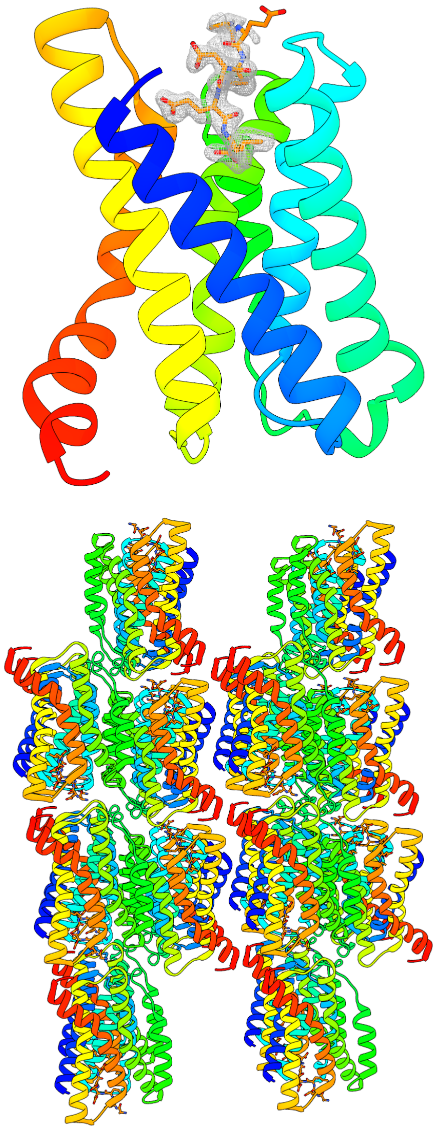
The KDELR formed crystals at two different pH conditions in the presence of TAEHDEL peptide. These crystals formed in the same LCP screen at 20 °C. The OMIT maps for the peptides in these crystals are seen in figure 4.7.

TAEHDEL pH 6.0

At a low pH, the crystals formed in 30% (v/v) PEG 550 MME, 100 mM MES pH 6.0, 100 mM ammonium phosphate dibasic. This structure closely resembles the TAEKDEL structure with an RMSD of 0.128 Å over 198 Cα. Crystal packing indicates a monomeric state. Data collection statistics can be seen in figure 4.5.

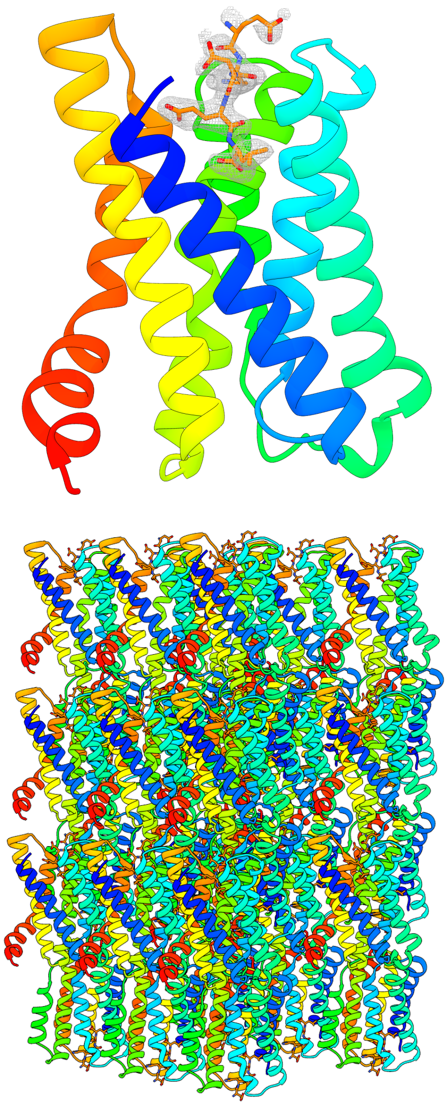
TAEHDEL pH 7.0

At a neural pH the KDELR-TAEHDEL crystals formed in 30% (v/v) PEG 500 DME 100 mM, HEPES pH 7.0, 100 mM potassium thiocyanate. This structure is globally similar to the other two peptide-bound structures, with an RMSD of 0.157 Å over 197 Cα to the TAEKDEL structure. The RMSD between the two TAEHDEL structures is 0.198 Å over 203 Cα. Crystal packing indicates a monomeric state, data collection statistics can be seen in figure 4.6.



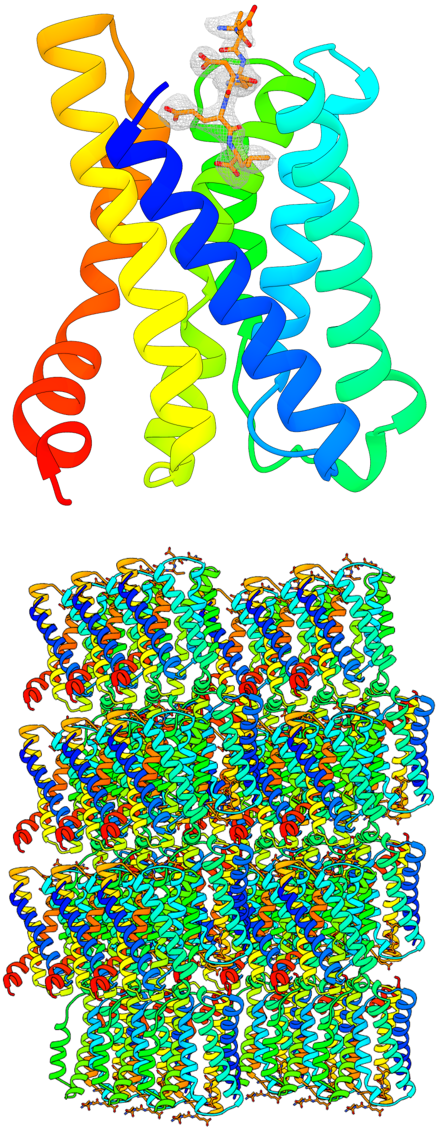
KDELR + TAEKDEL (6I6H)	
Data Collection	
Space Group	P 1 2_1 1
Cell Dimensions	
a, *b*, *c* (Å)	47.87, 37.50, 62.75
α, β, γ(°)	90, 95.42, 90
Wavelength (Å)	0.969
Resolution (Å)	47.66–2.00
	(2.05–2.00)
R_{pim}	0.096 (0.895)
R_{merge}	0.282 (2.588)
I/σI	6.8 (1.7)
Completeness (%)	99.2 (97.0)
Multiplicity	9.5 (9.1)
CC1/2	0.992 (0.721)
Wilson B-factor (Å²)	22.6
Refinement	
No. reflections	143906
Unique reflections	15154
R_{work}/R_{free}	0.19/0.23
No. atoms	
Protein	1755
(B_{res}), protein atoms (Å²)	25.5
(B_{res}), peptide atoms (Å²)	36.4
(B_{res}), waters (Å²)	44.3
(B_{res}), lipid atoms (Å²)	65.5
Ramachandran favoured (%)	97.6
Ramachandran outliers (%)	0.0
R.M.S deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.49

Figure 4.4 The Crystal structure of KDELR bound to TAEKDEL, polder OMIT map of the peptide is shown in grey mesh at 3σ. Crystal packing is shown below the asymmetric unit. On the right the data collection and refinement statistics for this structure can be seen.



KDELR + TAEHDEL pH 6	
Data Collection	
Space Group	P 1 2_1 1
Cell Dimensions	
a, *b*, *c* (Å)	47.62, 37.82, 62.42
α, β, γ(°)	90, 95.18, 90
Wavelength (Å)	0.969
Resolution (Å)	37.82–2.24
	(2.28–2.24)
R_{pim}	0.077 (0.739)
R_{merge}	0.111 (1.096)
$I/\sigma I$	6.5 (1.2)
Completeness (%)	97.1 (98.6)
Multiplicity	2.9 (2.9)
CC1/2	0.993 (0.457)
Wilson B-factor (Å²)	37.70
Refinement	
No. reflections	30374
Unique reflections	10578
R_{work}/R_{free}	0.24/0.27
No. atoms	
Protein	1730
(B_{res}), protein atoms (Å²)	47.35
(B_{res}), peptide atoms (Å²)	51.32
(B_{res}), waters (Å²)	52.83
(B_{res}), lipid atoms (Å²)	76.03
Ramachandran favoured (%)	98.6
Ramachandran outliers (%)	0.0
R.M.S deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.62

Figure 4.5 The Crystal structure of KDELR bound to TAEHDEL at pH 6, with a polder OMIT map shown in grey mesh at 2σ. Crystal packing is shown below the asymmetric unit. On the right the data collection and refinement statistics for this structure can be seen.



KDELR + TAEHDEL pH 7	
Data Collection	
Space Group	P 1 2_1 1
Cell Dimensions	
a, *b*, *c* (Å)	48.42, 38.2, 63.01
α, β, γ(°)	90, 95.43, 90
Wavelength (Å)	0.969
Resolution (Å)	36.58–2.41
	(2.45–2.41)
R_{pim}	0.135 (1.520)
R_{merge}	0.205 (2.331)
I/σI	4.6 (0.6)
Completeness (%)	98.9 (97.3)
Multiplicity	3.2 (3.3)
CC1/2	0.983 (0.276)
Wilson B-factor (Å²)	45.62
Refinement	
No. reflections	28402
Unique reflections	8986
R_{work}/R_{free}	0.22/0.26
No. atoms	1933
Protein	
(B_{res}), protein atoms (Å²)	55.50
(B_{res}), peptide atoms (Å²)	66.36
(B_{res}), waters (Å²)	55.84
(B_{res}), lipid atoms (Å²)	86.49
Ramachandran favoured (%)	98.1
Ramachandran outliers (%)	0.0
R.M.S deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.52

Figure 4.6 The Crystal structure of KDELR bound to TAEHDEL at pH 7, with a polder OMIT map of the peptide shown in gray at 2.5σ. crystal packing is shown below the asymmetric unit. On the right the data collection and refinement statistics for this structure can be seen.

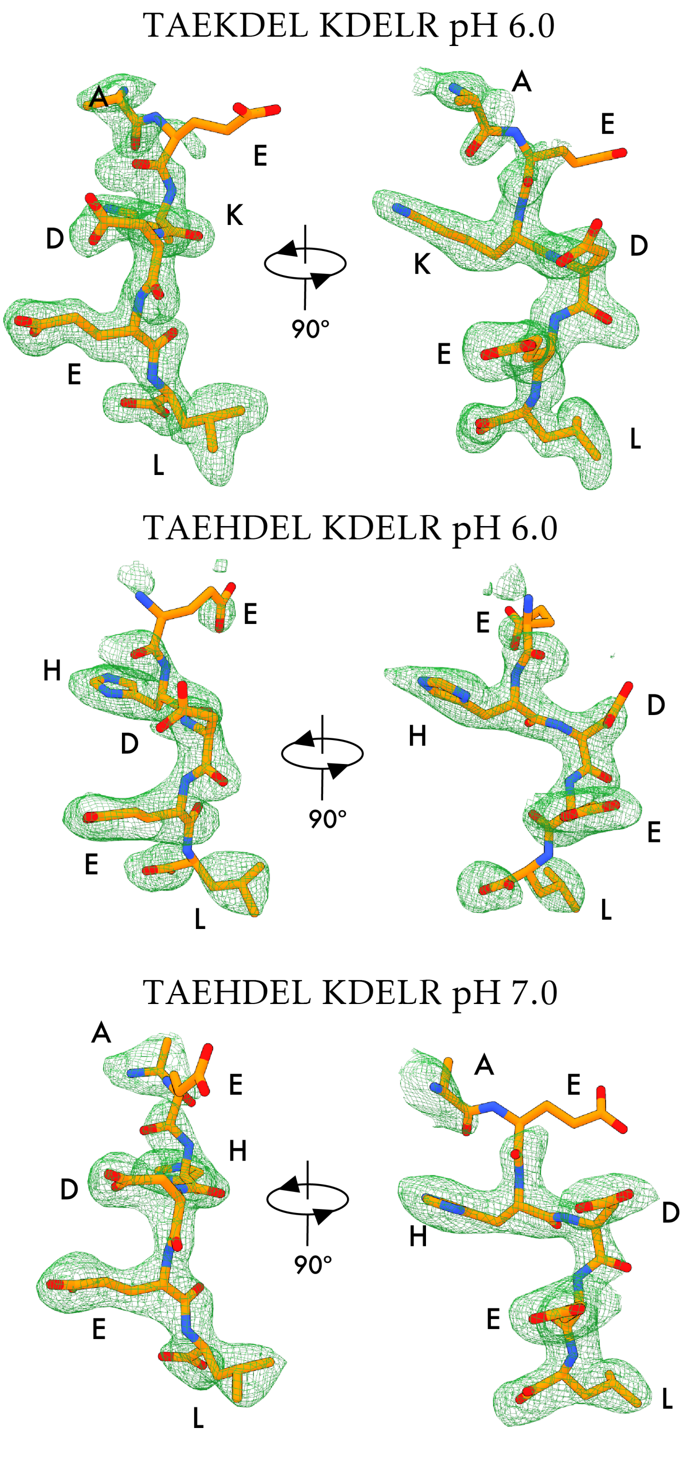


Figure 4.7 Polder OMIT maps of the bound peptides, TAEKDEL is shown at 3σ, TAEHDEL pH 6.0 at 2.4σ, and TAEHDEL pH 7 at 2.5σ (Liebschner et al., 2017).

4.4 Crystal Packing

All peptide-bound structures form the same crystal lattice, consisting of a plane of hydrophobic interactions around the TM regions of the KDELR, which are arranged in both vertical orientations relative to the plane (Peptide-binding site up and down - with respect to an imagined membrane). These planes then stack on-top of each other with a slight offset. Several interactions between these hydrophobic planes are shown in figure 4.8.

The apo structure packing is similar in composition, with a hydrophobic planes stacking at an offset, and the KDELR taking both vertical orientations. The distance between the planes is wider than the peptide-bound structures with no direct contacts between KDELR molecules in different planes. The packing within the planes also appears to have more spacing between KDELR molecules, this may be part of the reason why these crystals diffracted to a lower resolution.

In the sybody-bound structure, the characteristic hydrophobic plane remains. However, in-between the planes, Syb37 forms a scaffold with interactions to other sybodies and KDELR molecules from the adjacent plane, shown in figure 4.9. Effectively forming a sandwich of Syb37 between two layers of KDELR. While the sybody is evidently not required for KDELR crystallisation, here it performs its intended duty as a crystallisation scaffold. Presumably the conformational stabilisation and additional crystal contacts when compared to the apo structure allowed a higher resolution to be achieved.

The conformational differences of the KDELR in the bound and apo appear to lead to a different arrangement of the planes described above. Crucially helix 7 in the peptide-bound structures is involved in inter-plane crystal contacts that are not available to the apo conformation of the KDELR. These subtle conformational changes seem to alter the possible crystal contacts between the KDELR resulting in different crystal symmetries. The addition of a sybody opens

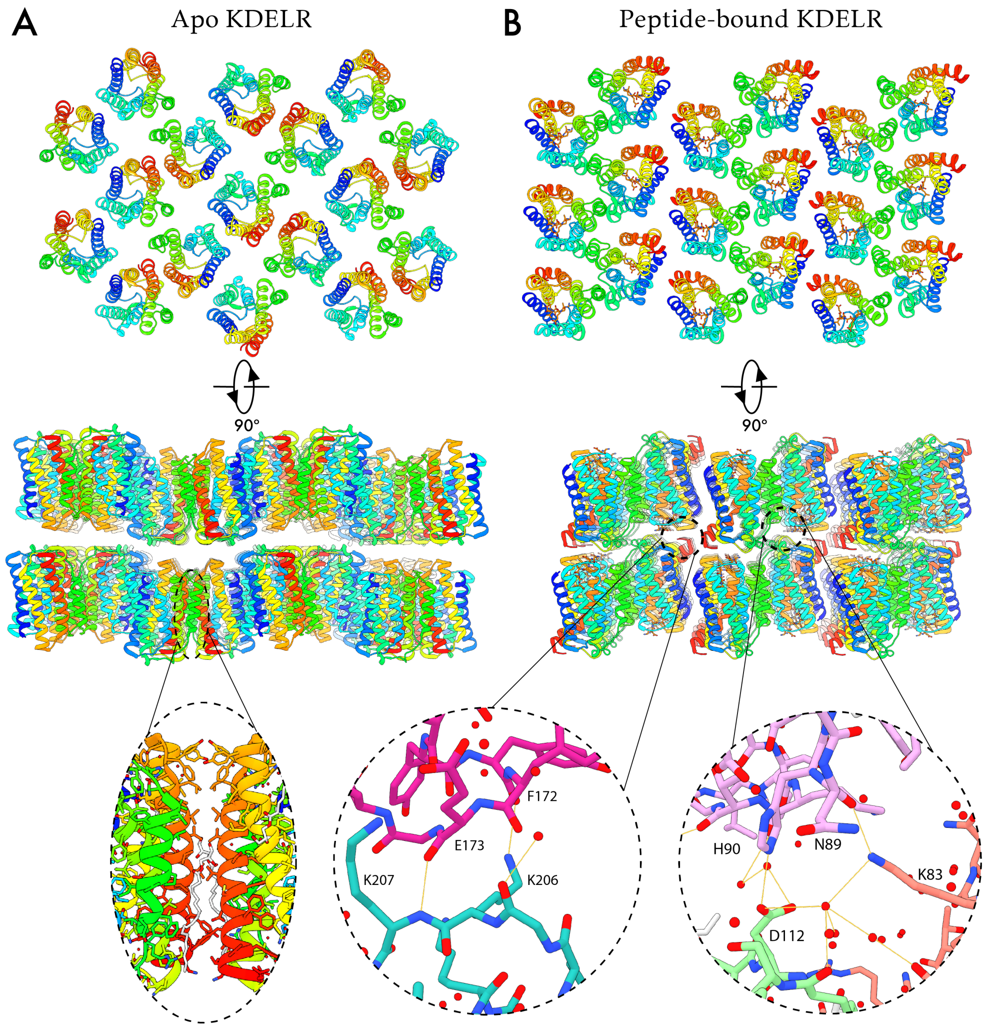


Figure 4.8 **A** Crystal packing of the apo KDELR structure, showing the planes of KDELR and how they pack. The inset shows a closer view of the hydrophobic packing between the planes of the KDELR. **B** The crystal packing of the peptide-bound structures, again showing the planes formed by the KDELR. The insets below show two regions where the planes interact directly, the different colours here represent different KDELR molecules.

up a new energetic landscape for crystal latices which, while following the same hydrophobic plane principle, leads to another crystal symmetry.

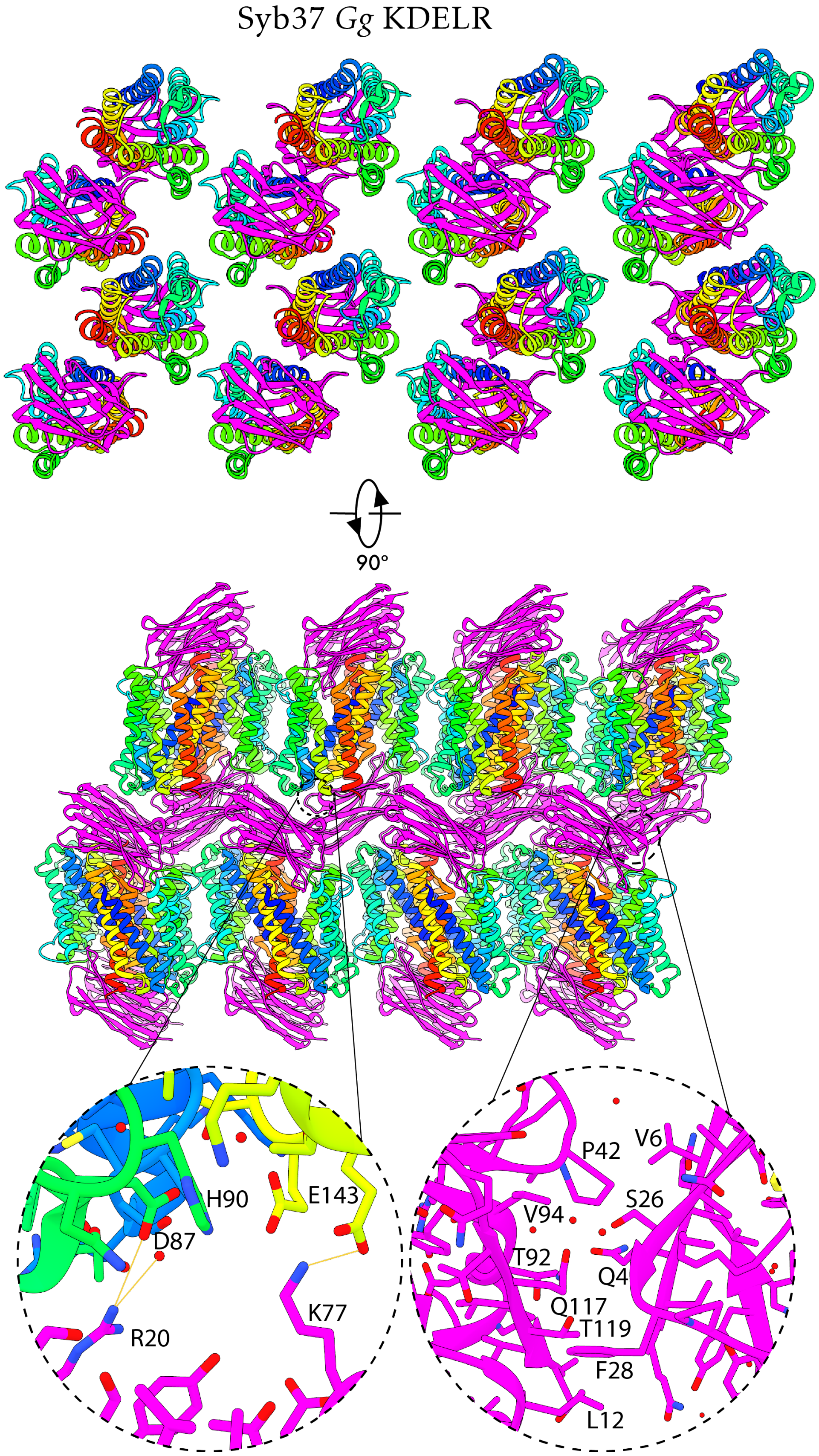


Figure 4.9 The crystal packing of the Syb37-KDELR structure, showing the sybody sandwiched between layers of KDELR. Packing interactions between Syb37 and KDELR as well as between two Syb37 molecules are shown in the insets.

4.5 Sample Preparation & Crystallisation

Gg KDELR was purified in detergent as described in the methods section 8.4.2, resulting in pure protein. The final SEC trace and SDS-PAGE of the *Gg* KDELR purification is shown in figure 4.10. In the case of peptide-bound structures, peptide was added to concentrated KDELR from a high concentration peptide stock before LCP preparation. KDELR-Syb37 crystals were made with a complex of the two purified components, which was polished by a SEC step, the final traces and SDS-PAGE are seen in figure 4.11 and 4.12.

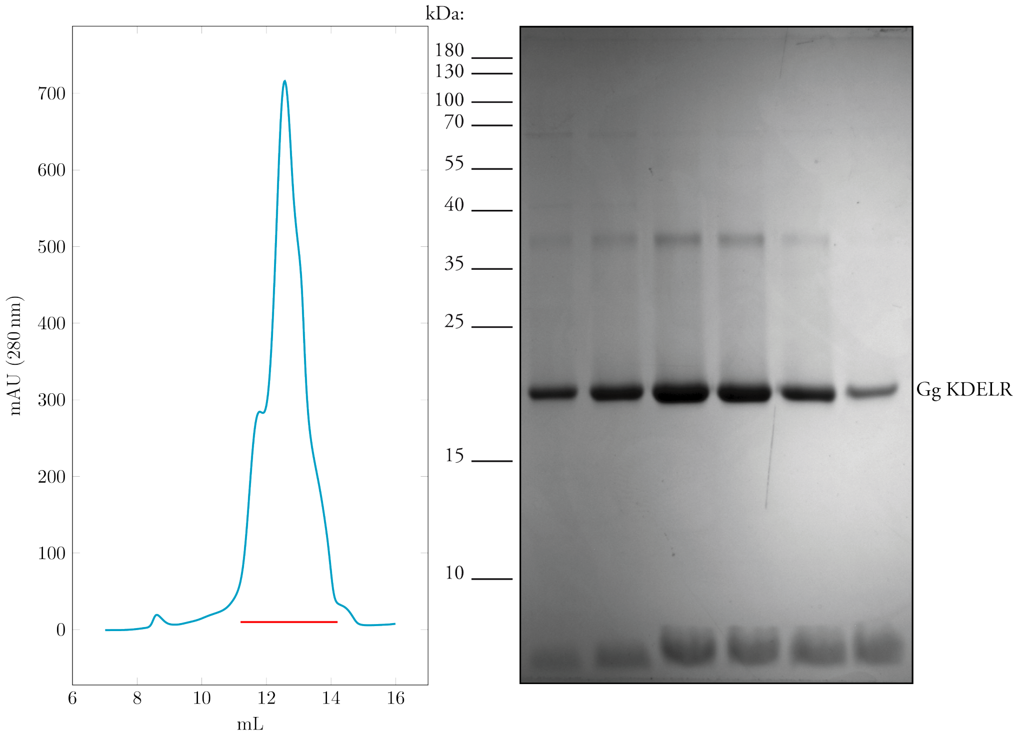


Figure 4.10 SEC S200 trace and SDS-PAGE of a typical *Gg* KDELR purification, the red line indicates the region for which samples were taken and run on the gel.

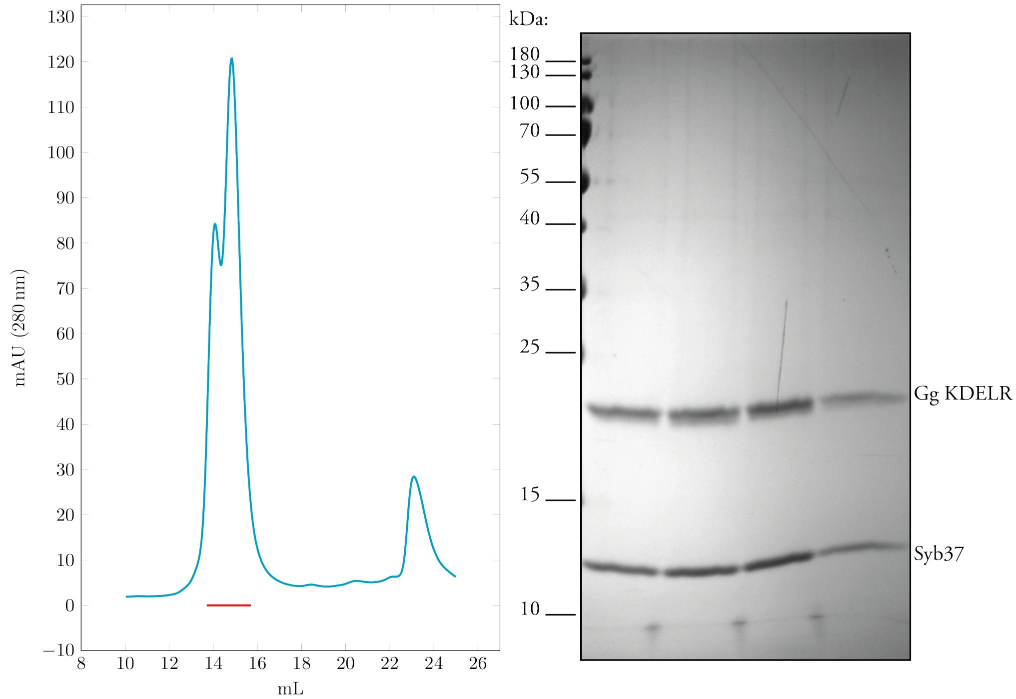


Figure 4.11 SEC SRT-C SEC 300 trace and SDS-PAGE of *Gg* KDELR-Syb37 complex, the red line indicates the region for which samples were taken and run on the gel.

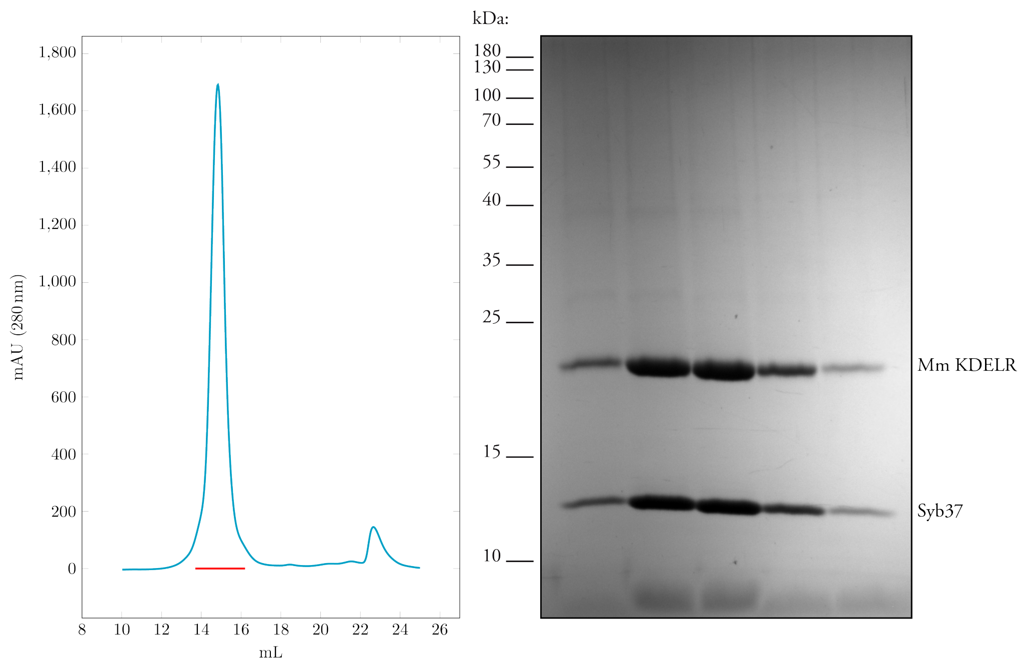


Figure 4.12 SEC SRT-C 300 trace and SDS-PAGE of *Mm* KDELR-Syb37 complex, the red line indicates the region for which samples were taken and run on the gel.

4.5.1 Crystallisation

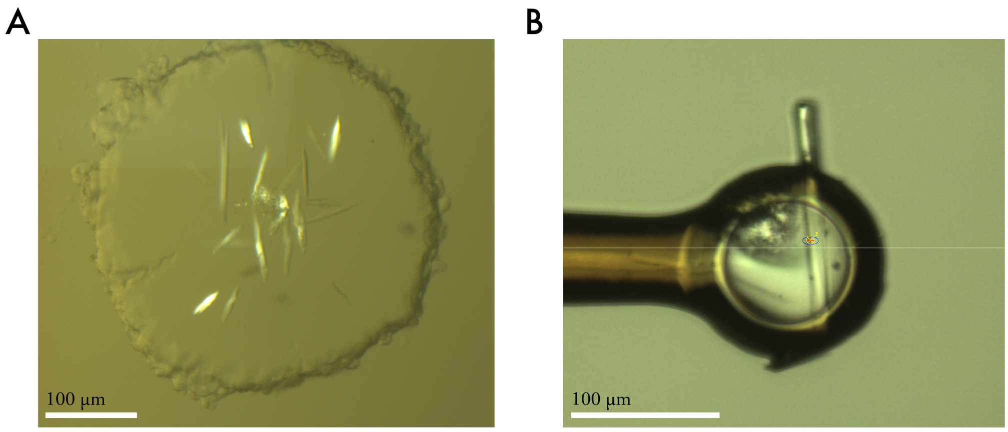


Figure 4.13 Example of crystals obtained. **A** KDELR-SYB37 crystals in the drop. **B** KDELR TAEKDEL crystal in a loop.

Crystallisation was initially attempted by both vapour diffusion and LCP methods. KDELR precipitates rapidly in most vapour diffusion conditions tested, even at very low concentrations. LCP crystallisation conditions were screened at 4 °C and 20 °C, in monoolein (MO), using a custom crystallisation screen, covering a range of polyethylene glycols (PEGs), pH, buffer and salt conditions. The crystals were harvested directly from the screens, some can be seen in figure 4.13. No optimisation was required, the mean resolution of crystals was ~2.5 Å (with the exception of *Mm* KDELR-Syb37), dozens of crystals were harvested and screened at a synchrotron.

4.6 Data Collection

All of the crystal structures solved came from crystals that formed in LCP. Details of the conditions and setup can be found above and in the methods section 8.6. All data, except *Mm* KDELR-Syb37 were collected at the I24 microfocus beam line at the Diamond Light Source, Didcot UK. *Mm* KDELR-Syb37 data was collected at the Proxima2A beam line at the Soleil Synchrotron, Gif-sur-Yvette, France.

	Apo	Syb37	Syb37	TAEKDEL (pH 6)	TAEHDEL (pH 6)	TAEHDEL (pH 7.0)
	Gg KDELR	*Mm* KDELR	*Gg* KDELR	*Gg* KDELR	*Gg* KDELR	*Gg* KDELR
PDB:	6I6B		6I6J	6I6H		
Space group	C 2 2 2_1	P 2_1 2_1 2	P 2_1 2_1 2_1	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1
Beamline	I24	Proxima2A	I24	I24	I24	I24
Wavelength (Å)	0.9686	1.0055	0.968	0.9686	0.9686	0.9686
Φ (°)	180	360	180	180	180	180
ΔΦ (°)	0.5	0.1	0.2	0.5	0.2	0.2
Transmission (%)	10.51	20.00	14.50	10.09	41.27	41.27
Beamsize (µm)	9x6	5x10	9x6	9x6	7x7	7x7
Exposure (s)	0.05	0.025	0.1	0.05	0.01	0.01

Table 4.1 Data collection statistics for all structures of the KDELR.

In all cases at least 180° of data were collected, this ensured that full completeness could be achieved for any space group as well as reasonable multiplicity while avoiding radiation damage. As the crystals were very small and often not visible in the loop, raster scans were performed to locate and centre the crystals in the loops. Small beam sizes were used to ensure most of the beam would enter the crystal at all times.

4.7 Data Processing

4.7.1 Indexing, Scaling & Merging

Diffraction data collected at Diamond was all processed with the Xia2 pipeline. With the exception of the TAEHDEL KDELR structures, all with the dials pipeline, which indexes, integrates and merges with DIALS and scales with AIMLESS (**Winter, 2010; Winter et al., 2018; Evans & Murshudov, 2013**). The TAEHDEL KDELR data showed better statistics when run through XDS and XSCALE (**Kabsch, 2010**). *Mm* KDELR-Syb37 data was indexed and integrated with XDS, scaled and merged in AIMLESS (**Kabsch, 2010; Evans & Murshudov, 2013; Evans, 2011**).

4.7.2 Phasing & Molecular Replacement

Experimental Phasing

Before KDELR-Syb37 crystals were grown and observed to diffract, thus enabling MR with a sybody model, attempts were made to experimentally determine the phases of the *Gg* KDELR structures. Two approaches involving heavy metal derivatisation were tested. Firstly, prederivatisation, followed by a SEC run to remove excess metals to simplify finding the anomalous signal of bound metal atoms, and secondly, heavy atom soaking. Selenomethionine incorporation was considered but this is thought to be difficult in yeast expression systems, so was not attempted (**Walden, 2010**).

It has been observed that alpha helical membrane proteins, which have been experimentally phased with heavy atom salts, were predominantly solved using mercury salts, followed by platinum, lead and gold. The most popular compound being methyl mercury chloride (MMC) (**Parker & Newstead, 2013**). The most common data collection strategy was single wavelength anomalous dispersion (SAD) (**Parker & Newstead, 2013**). Thus focus was placed on these heavy atom compounds.

Initially, to screen compounds for suitability, Fluorescence Detection of Heavy Atom Labelling was used to screen two popular mercury compounds, MMC and $HgCl_2$ (**Chaptal et al., 2010**). In this method, protein is incubated with the heavy atoms of choice, if this derivatises the protein via free cystienes, the dye tetramethylrhodamine-5-maleimide (TMRM) will no longer be able to covalently bind these. Thus if fluorescence is abolished with the addition of a heavy atom, this indicates a possible derivatisation (**Chaptal et al., 2010**). Of the two mercury compounds tested, $HgCl_2$ appeared to bind KDELR covalently (figure 4.14), unfortunately it also appears to cause complete aggregation of the KDELR.

To test the effect of heavy atom derivatisation on the stability of the KDELR, protein was incubated and loaded onto an S200 column. Several heavy atom

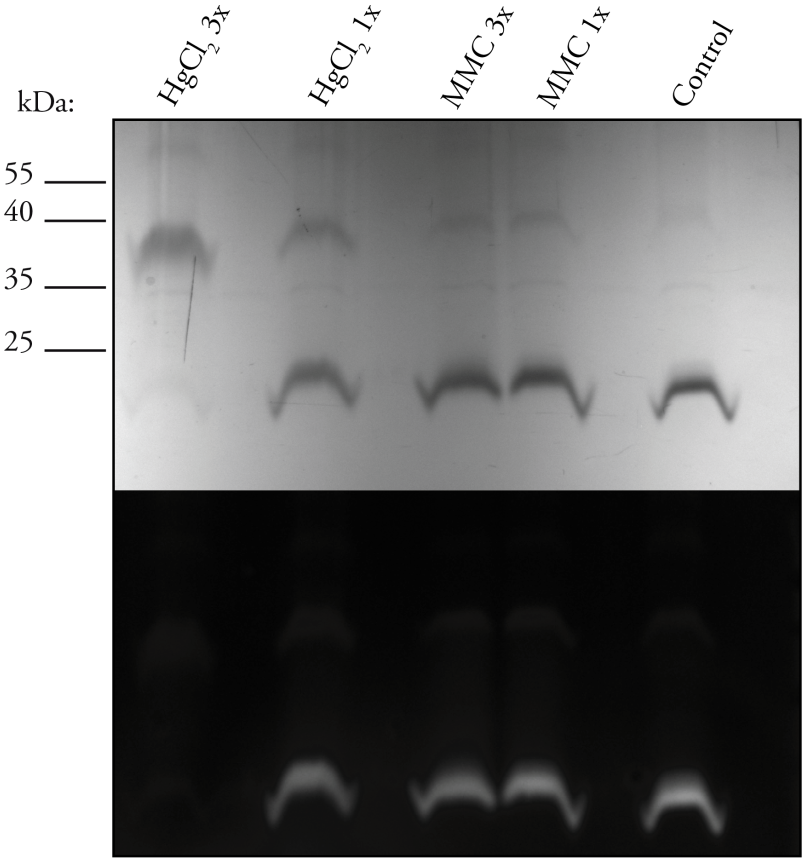


Figure 4.14 FD-HAL of *Gg* KDELR with two mercury compounds, only a 3x molar excess of $HgCl_2$ fully derivatises the KDELR and appears to induce oligomerisation.

incubations appear to be tolerated very well (figure 4.15). *Gg* KDELR that was prederivatised with MMC, run down an SEC, crystallised and diffracted, but no anomalous heavy atom substructure was observed. Datasets were collected, optimised for a SAD experiment, with the wavelength set to maximise anomalous signal from the heavy atom, which was determined experimentally on the beam line with X-ray fluorescence, as well as collecting for high multiplicity and completeness in order to observe the small differences in the Bijvoet pairs resulting from the presence of anomalous scatterers.

Thus crystals were soaked with heavy metal compounds after they had formed, screening for a variety of compounds and soaking times, summarised in table 4.2.

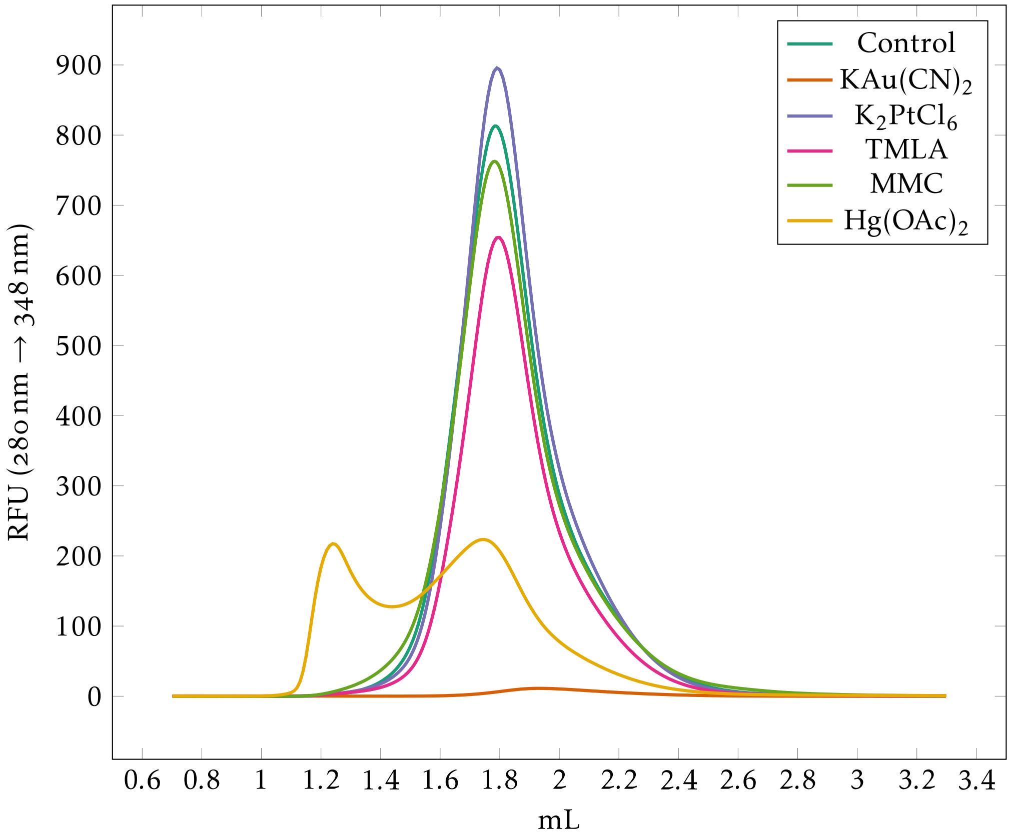


Figure 4.15 Screen of various heavy atom salts incubated with *Gg* KDELR and run down an S200 10:50 column. Column was not calibrated with molecular weight standards.

None of the crystals that were soaked showed any anomalous heavy atom substructure. Most crystals did not dissolve during the soaking process, though many did not diffract, or did so only poorly. Those that did, had datasets collected as above, but no phases could be determined.

Compound	Concentration (mM)	Time (h)
MMC	0.5	4 & 6
MMC	0.4	20
$HgCl_2$	1	20
$Hg(OAc)_2$	5	4 & 17
$Hg(OAc)_2$	0.5	17
$KAu(CN)_2$	0.5, 1 & 5	17
K_2PtCl_6	10	17

Table 4.2 List of all the heavy atom compounds used for soaking TAEKDEL-KDELR crystals.

Molecular Replacement

`Phenix.phaser` is a maximum likelihood MR program which, like a Patterson search, is composed into a rotation function (RF) followed by a translation function (TF) (McCoy et al., 2007). Maximum likelihood models errors in the data, while Patterson methods do not. In fact, the Patterson rotation function is essentially the first term of a Taylor series expansion of the likelihood rotation function (Evans & McCoy, 2008). Due to the central limit theorem, reciprocal structure factor distributions may be approximated as a Gaussian distribution (McCoy, 2004). Each molecule needs six parameters to define its orientation, three rotational angles and three translations, however, performing a 6N dimensional (for N number of molecules in the asymmetric unit) search for a molecule is computationally expensive (Evans & McCoy, 2008). The search is thus split into the two functions mentioned above. The translational search effectively has the same target function as a refinement. While errors are modelled for phased structure factors $\mathbf{F_o}$ (observed structure factor) and $\mathbf{F_c}$ (calculated structure factor), the phases are a nuisance variable and are integrated out because we only observe the structure factor amplitude F_o. The structure factor of a model in the unit cell is calculated for each translation and scored according the target likelihood function for a reflection, (equation 4.1), a two dimensional Gaussian with a variance dependent on σ_Δ^2 and σ_F the errors in atomic positions and scattering, and the experimental errors respectively, forming a Rice distribution see McCoy (2004) for details.

$$P-TF_r = \frac{2F_o}{\sigma_\Delta^2+\sigma_F^2} exp\left(-\frac{F_o^2+D^2F_c^2}{\sigma_\Delta^2+\sigma_F^2}\right) I_0\left(\frac{2F_oDF_c}{\sigma_\Delta^2+\sigma_F^2}\right) \tag{4.1}$$

During the rotational search no structure factors can be calculated, as only the orientation is known, however atomic structure factors can be added with relative phases to calculate F_m, the amplitude of the model structure-factor contribution. This is performed for all symmetry relatives of the model $\{F_m\}_{sym}$. A phase variable is introduced between F_o and the symmetry relative of F_m with the largest amplitude F_{big}, the remaining $\{F_m\}_{sym}$ relatives remain unphased and result in a random walk distribution around F_{big}. As above, the phases must integrated out, placing a phase on F_{big} results in the smallest probability distribution of $\{F_m\}_{sym}$. The likelihood function for a reflection is thus equation 4.2, a two dimensional Gaussian with with a variance Σ_S related to the random walk of the smaller $\{F_m\}_{sym}$, forming a Rice distribution (McCoy, 2004).

$$P-RF_r = \frac{2F_o}{\Sigma_s} exp\left(-\frac{F_o^2+D^2F_{big}^2}{\Sigma_S}\right) I_0\left(\frac{2F_0DF_{big}}{\Sigma_S}\right) \tag{4.2}$$

`Phenix.phaser` attempts to maximise the log-likelihood gain (LLG) for a model placed in the unit cell and also reports a translation-function Z-score (TFZ), the number of standard deviations over the mean (Z-score) for the LLG, generally a LLG over 60 and a TFZ over 8 indicate a correct solution (Oeffner et al., 2018; McCoy, 2017). Another indication of a solution is an R factor ($\Sigma||F_o| - |F_c|| / \Sigma|F_o|$) around 45–50%, with the principal indication being maps that show new, plausible information not present in the model (McCoy et al., 2007).

MR search models generally require at least 35% sequence identity to the target, corresponding to a C_α RMSD of about 1.5 Å, to find a solution (Abergel, 2013).

The sequence identity is only important in as much as it allows an estimate of RMSD, given this can only be calculated after successful MR. The RMSD requirements are variable and can depend on model completeness, i.e. the fraction of scattering, and the number of reflections. The various parameters affecting MR can compensate for weaknesses in others (McCoy, 2017).

The closest homologue to the *Gg* KDELR with a structure, is *Os* SWEET2b (PDB: 5CTG), with a sequence identity of only 25%. While the structure was used for MR attempts, as well as homology models, this was not successful. Upon collecting data of KDELR-Syb37 data, it became possible to attempt MR using the scaffold structure, on which the sybody was based, as a search model. The success of this hinges on the fact that Syb37 is 128 residues and about 14.4 kDa while the KDELR has 212 residues with about 25.3 kDa, thus making about 36% of the structure. As only the CDR regions of the Syb37 are different from the scaffold structure, there is a chance that using a this as a search model would be successful.

Mm KDELR-Syb37 Molecular Replacement

MR using `phenix.phaser` with the search model being the scaffold nanobody structure on which Syb37 was based (PDB: 5M13), was performed. This resulted in a solution with a LLG of 89 and TFZ of 11.3. After one round of refinement with REFMAC, the R_{Work} was 0.48 and the R_{free} 0.52 (Murshudov et al., 2011). The maps resulting from this can be seen in figure 4.16. Density could be seen for a helical protein, though it was very weak in places. Initially backbones of helices were modelled into the densities with O (Jones et al., 1991). As the maps improved, seven helices could be successfully placed into the density, however no density for side chains in the KDELR could be observed. Fortuitously at this point, higher diffracting *Gg* KDELR-Syb37 crystals were identified which showed unambiguous side chain density. Therefore the *Mm* KDELR-Syb37 structure was abandoned. Later, the model obtained from the *Gg* KDELR-Syb37 data was used as a model in the *Mm* data. The complete model brought out some side chain density in the core of the protein, though most residues could not have any rotamers assigned.

Gg KDELR Molecular Replacement

Using the model of the KDELR generated from the *Mm* KDELR-Syb37 structure, MR was performed on the *Gg* KDELR-Syb37 structure, with an LLG of 111.9 and TFZ of 12.0. After this the poly-ala helices were extended and the connectivity improved. This improved the density and `phenix.autobuild` was run, resulting in a high quality model, which was subsequently refined and used as a MR model for the remaining structures, resulting in clear, unambiguous solutions (Terwilliger et al., 2008).

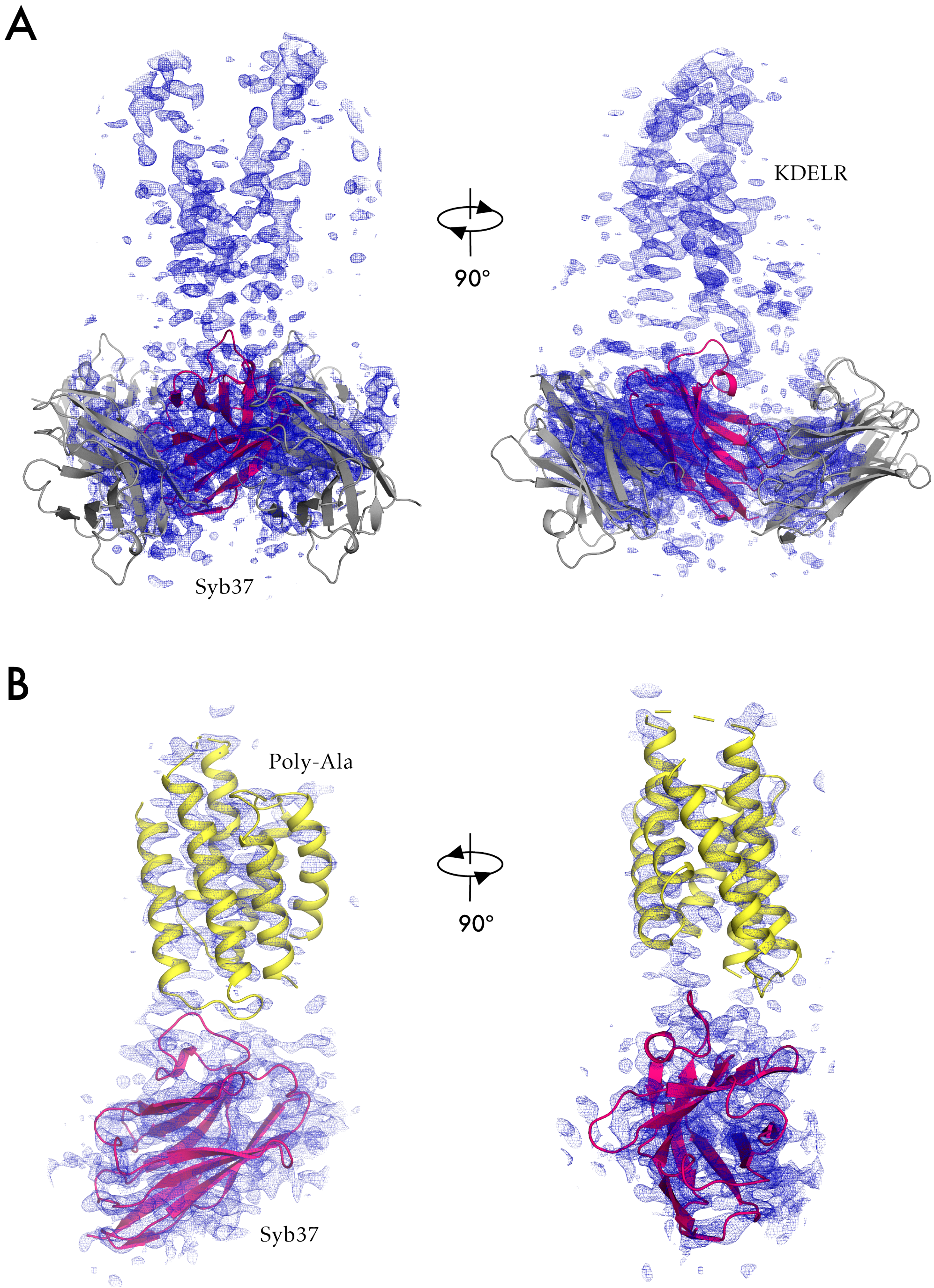


Figure 4.16 **A** $2mF_o - DF_c$ map of *Mm* KDELR-Syb37 phaser solution after one round of refinement with REFMAC. Symmetry mates of the sybody are shown in grey. **B** The same $2mFo - DFc$ map now shown with the final *Mm* KDELR poly-ala model used for MR on the *Gg* KDELR-Syb37 structure. Map contoured at 1.5σ.

4.7.3 Model building & Refinement

Refinement was performed using Coot for manual model modification and validation, and `phenix.refine` and BUSTER for automated maximum likelihood refinement procedures (**Emsley et al., 2010; Adams et al., 2010; Afonine et al., 2012; Bricogne et al., 2017**). MR with maximum likelihood attempts to maximise the probability of the model, given the observations, taking into account errors and chemical restraints regarding bond lengths and angles, etc. Effectively optimising $P - \text{refinement} = P - \text{chemistry} \cdot P - \text{Xray}$ where $P - \text{Xray}$ is the same as the translation function 4.1 for MR. Refinement invokes Bayes theorem, with $P - \text{chemistry}$ as the prior and with $P - \text{Xray}$ being the likelihood or $P(\text{data;model})$, with $P - \text{refinement} = P(\text{model;data})$ (**McCoy, 2004**). In all cases, refinement was run with Translation/Libration/Screw (TLS) refinement, which describes the local movement of atoms in a molecule by assuming that each atom is part of a rigid body that is displaced normally about a mean position (**Urzhumtsev et al., 2011; Schomaker & Trueblood, 1968**). TAEHDEL structures were refined with Local Structure Similarity Restraints (LSSR) targets from the higher resolution TAEKDEL structure.

To determine when refinement was complete, several parameters were observed. The R_{work} and R_{free} values were stabilised and could no no longer be improved, the bond distance and angle deviations were minimal, there were few if any Ramachandran outliers, clash scores were minimal and no more water or lipid molecules could reasonably be modelled to explain non-protein density. These parameters were checked manually in the maps, in the direct output of the refinement program, or calculated with Molprobity (**Williams et al., 2018**). The refinement statistics are shown in the tables above, examples of the final maps are shown in figure 4.17 and the Ramachandran plots are in figure 4.18.

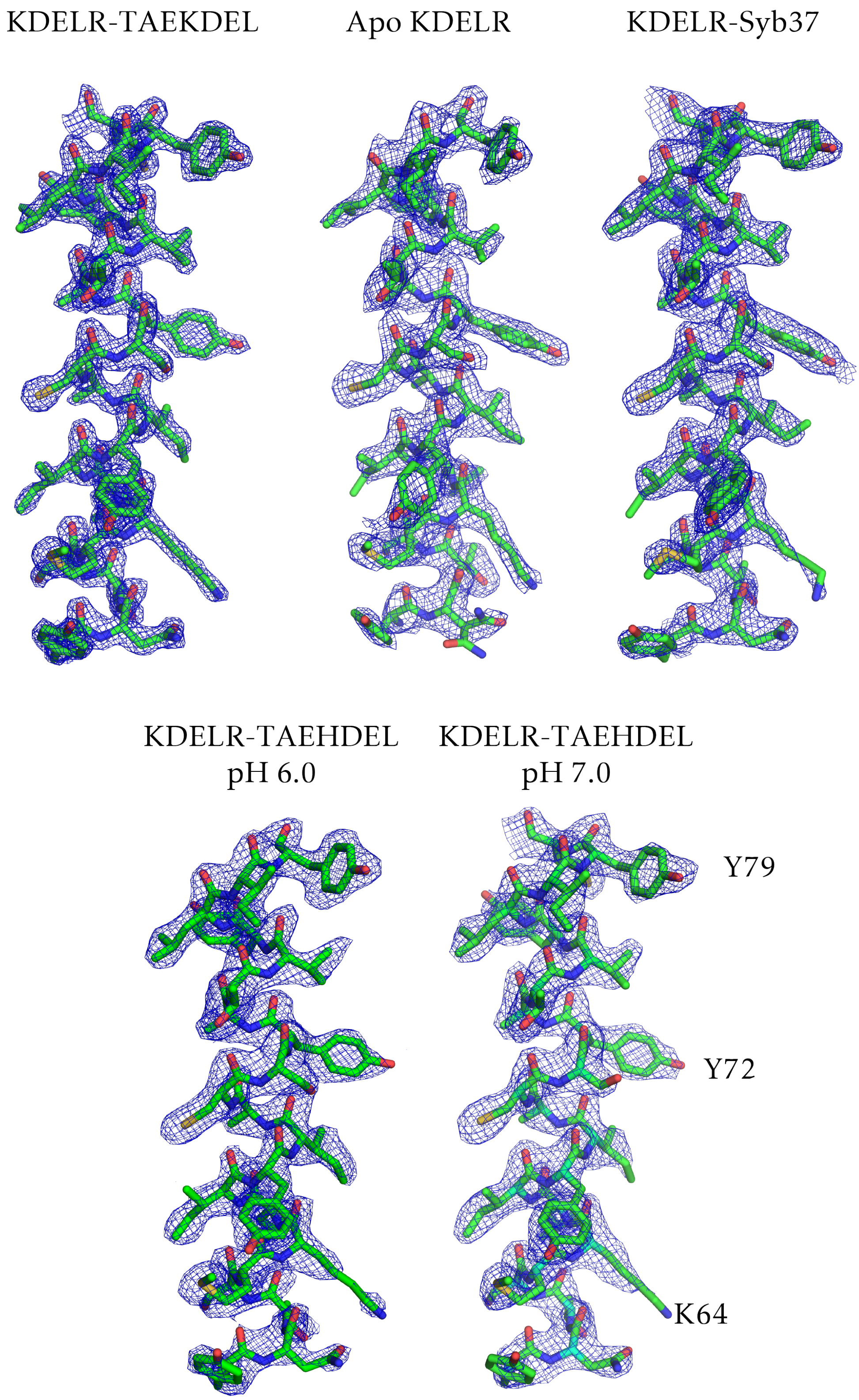


Figure 4.17 Electron density around helix three of the KDELR in all of the structures, $2mF_o - DF_c$ map contoured at 2σ for the TAEKDEL structure and 1.5σ for all others.

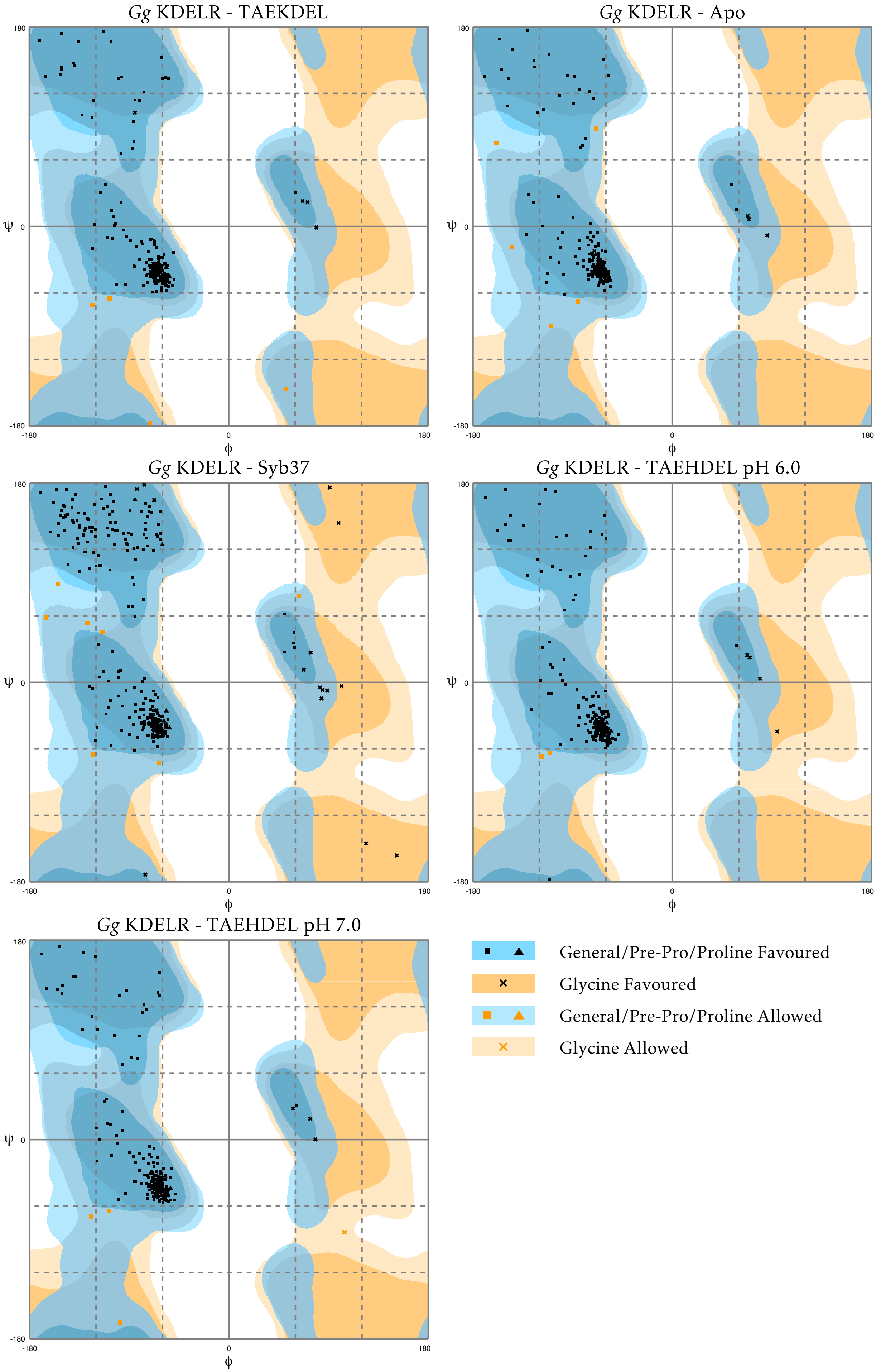


Figure 4.18 Ramachandran plot for all refined crystal structures, calculated with RAMPAGE (Lovell et al., 2003).

Peptide

In all three peptide-bound structures, the peptide was present with high occupancy in the same binding site. At least the four C-terminal residues (H/KDEL) were always visible in the density and well resolved. In some cases the -5 and -6 residues could be modelled, though the density here was noticeably weaker, presumably due to greater conformational flexibility of the residues outside of the binding pocket of the KDELR. In figure 4.7 polder OMIT maps are shown, with clear density for the modelled peptides.

Short Hydrogen Bond

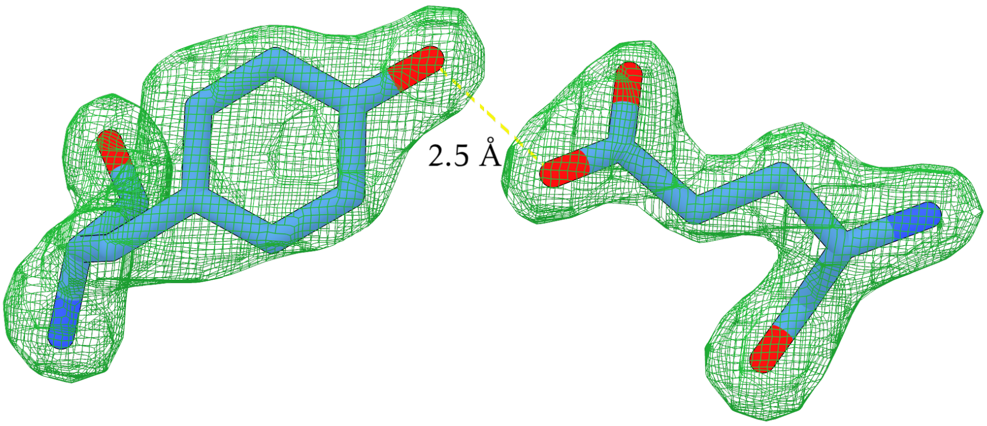


Figure 4.19 Polder OMIT map of E127 and Y158 in KDELR-TAEKDEL structure, at 5σ (Liebschner et al., 2017).

During the refinement process of the TAEKDEL bound structure, a short hydrogen bond between E127 and Y158 appeared in the density, however `phenix.refine` would flip the Y158 rotamer out of this configuration, creating strong positive difference density. A PDB LINK card between the two residues was introduced and refinement was moved to BUSTER. This resolved the flipping of Y158 and the difference density was no longer observed. To ensure the two oxygen atoms were being modelled with the correct distance, Polder OMIT maps were calculated for the two residues and can be seen in figure 4.19. Later it became clear that the LINK card was not required for correct refinement in BUSTER and this was removed for refinement of the TAEHDEL structures.

Hydrogen bonds between in protein structures are generally between 2.5–3.2 Å, averaging 3 Å (Jeffrey, 1997). A search of the PDB, looking for high resolution structures containing short hydrogen bonds by Qi & Kulik (2019) with donor-acceptor lengths shorter than 85% of the van der Waals radii of the two heavy atoms, thus mostly H-bonds shorter than 2.6 Å, found that Tyr is frequently involved in short hydrogen bonds, often with Asp or Glu residues. Short non-covalent contacts appeared most stable in the 2.4–2.8 Å range (Qi & Kulik, 2019). The precise position of the hydrogen between these two residues cannot be observed with the X-ray data, and would probably necessitate a neutron scattering experiment that is not feasible with LCP crystals, or an Nuclear Magnetic Resonance (NMR) experiment. Assuming Y158 is the proton donor, the hydrogen bond angle between acceptor-donor-donor antecedent is 107.3° and the angle between acceptor antecedent-acceptor-donor is 113.3° falling close to ideal trigonal or tetrahedral values (Stickle et al., 1992).

4.8 Conclusion

Data was collected for six different KDELR structures by X-ray crystallography, five of which could be used to refine atomic resolution maps. The solution of the phase problem hinged on the use of the sybody scaffold as an MR search model, as heavy atom phasing was proving challenging. The original purpose of the sybody was to enable crystallisation of the KDELR, as this was proving difficult with *Mm* KDELR. Indeed *Mm* KDELR only crystallised in the presence of Syb37, the *Gg* KDELR homologue, which was discovered after the Syb37 had been generated, was able to form crystals in its absence. Nevertheless the Syb37-KDELR structure proved to be the crucial one, as the phasing of this structure allowed the remaining structures to be solved with relative ease.

The use of the *Gg* KDELR homologue was key in obtaining high resolution information, highlighting the importance of effective homologue screening and characterisation. Despite high degrees of similarity, *Mm* and *Gg* KDELR show very different characteristics during crystallisation. Due to the high resolution afforded by the *Gg* KDELR crystals, initial model building and refinement could largely be automated, only requiring manual intervention in the final polishing stages.

Analysis of the structural features of the KDELR and a comparison between the KDELR structures and other related proteins will follow in chapter 5 Structural Analysis of the KDELR.

5 Structural Analysis of the KDELR

The crystal structure of the KDELR was successfully solved with LCP crystals in a variety of states. The five different structures have been briefly outlined in chapter 4 and will be analysed and compared here. Due to the high resolution of the structures it was possible to identify the KDEL-peptide binding region of the KDELR without complication. Comparisons between the apo and peptide-bound structures allowed for the identification of a putative pH sensing mechanism as well as the identification of possible COPI and COPII binding surfaces of the KDELR.

First the general structural properties of the KDELR will be described, then the various interactions between KDELR and Syb37 and the KDEL peptides will be analysed, compared and contrasted. Following this, mechanistic hypotheses regarding the pH-dependence and conformational changes allowing the communication of the state of the luminal side of the KDELR to the cellular COPI and COPII machinery will be developed.

5.1 Topology & Surface Properties

The topology of the KDELR, seen in figure 5.1, shows seven transmembrane helices arranged in a 3-1-3 topology, predicted from sequence analysis. There are two pseudo symmetric THBs displaying a 1-3-2 arrangement of helices, connected by a linking helix (4). The two THBs can be overlaid with an RMSD of 2.54 Å over 72 Cα atoms.

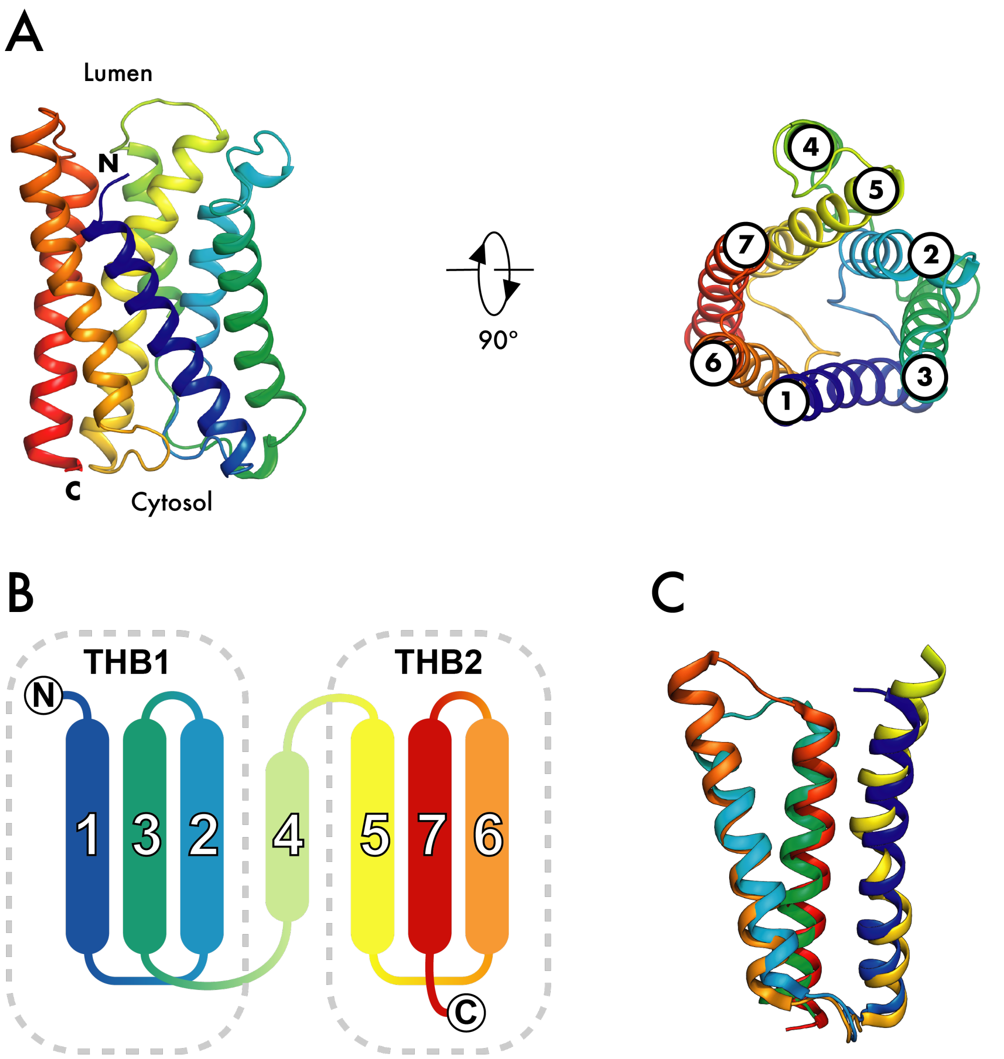


Figure 5.1 **A** Cartoon representation of the KDELR in the apo state, coloured blue to red from N to C-terminus, with helix numbers indicated in the 90° rotation to the right. **B** Topology diagram of the KDELR, with triple helical bundles (THBs) indicated. **C** The two THBs superimposed.

The surface of the apo KDELR displays three features of note, all shown in figure 5.2. Firstly there is a clear, approximately 27 Å wide, hydrophobic membrane region. The width of this membrane spanning region is short, but in agreement with other known membrane proteins resident in the ER and Golgi (**Sharpe et al.**, 2010; **Parker & Newstead**, 2017). Short TM domains are thought to be a possible mechanism for Golgi retention (**Banfield**, 2011; **Cosson et al.**, 2013). Secondly, we observe a large polar cavity on the predicted luminal side of the KDELR, which is where the KDEL-motif would be recognised. The cavity is flanked by side chains from helix 1 to 3 from the N-terminal THB and helix 5 to

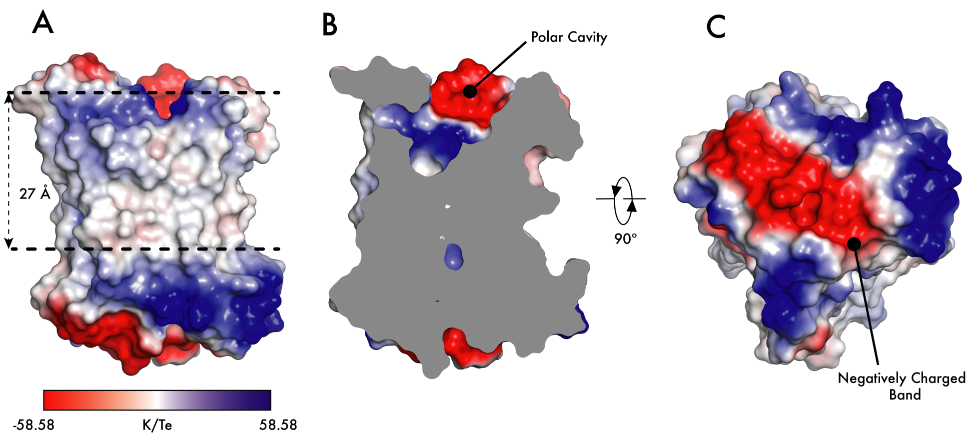


Figure 5.2 Electrostatic surface representation of the apo KDELR. **A** The placement of the receptor within the membrane. **B** The polar cavity on the luminal side with a sliced volume representation. **C** the negatively charged band on the cytosolic face.

7 from the C-terminal THB, and measures 13 Å x 15 Å x 12 Å. The electrostatic surface is charged, with a pronounced dipolar character contributed by R5 (helix 1) and R169 (helix 6), which are positioned opposite E117 (helix 5) and D177 (helix 7). Finally, on the cytosolic face, which unlike the luminal side, projects out from the membrane, a prominent band of negative charge runs down its centre. The negative charge is contributed by several acidic residues conserved in the mammalian KDELRs: D87, E143, E145, and the C terminus of helix 7.

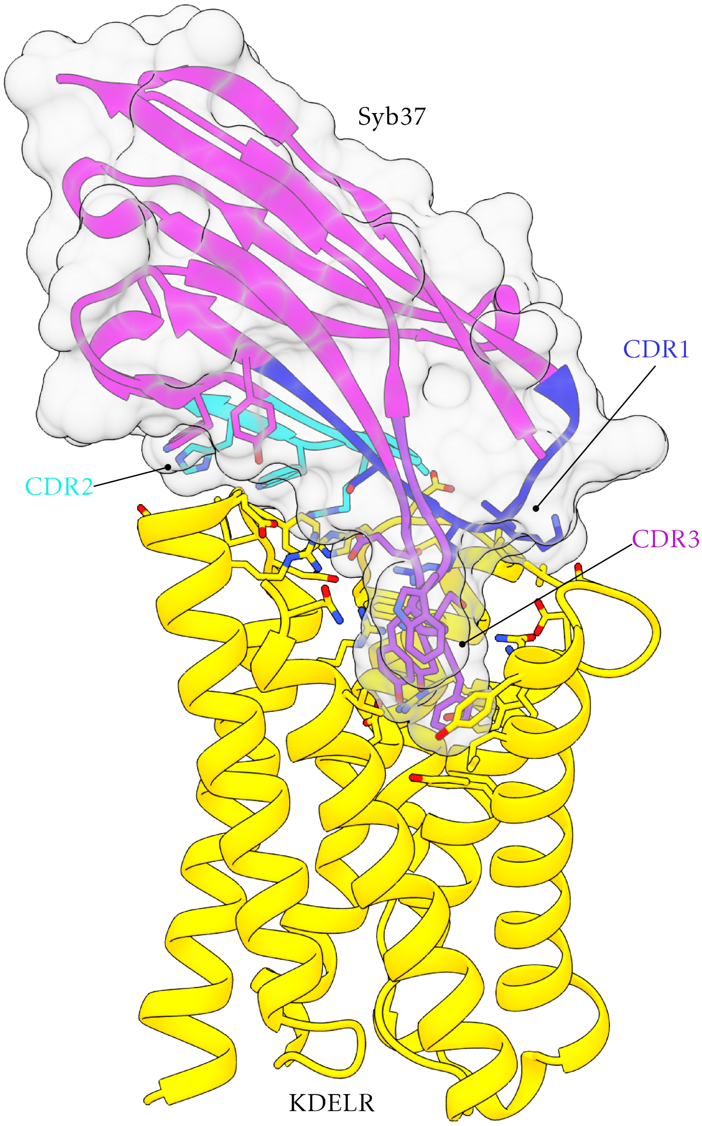


Figure 5.3 Crystal structure of Syb37-KDELR complex, with the surface of Syb37 shown in transparent grey. CDR1 is displayed in blue, CDR2 in cyan and CDR3 in purple. Residues involved in the interaction surface are shown in stick representation.

5.2 Sybody and Peptide Binding

The luminal face of the KDELR presents the polar cavity shown in figure 5.2. This cavity is where both the Syb37 and TAEKDEL peptide bind the receptor. Both of these interactions will be described below. While both interactions engage with the polar cavity, the nature of the interactions remains fundamentally different. Syb37, given its much larger size, has more opportunity to make a greater number of interactions, which are more hydrophobic in nature, while the TAEKDEL peptide has a larger proportion of hydrogen bonding and ionic interactions. TAEHDEL peptides bind in much the same way was the TAEKDEL peptide.

5.2.1 Sybody Binding

Syb37 and KDELR form a tight interaction with a surface area of 1202.9 Å^2, as determined by PISA (**Krissinel & Henrick**, 2007). In figure 5.3 one can see that CDR3 enters the polar cavity where it forms the majority of the interactions between Sybody and KDELR. Other parts of Syb37 do form interactions with the KDELR and an overview of the interactions is given in figure 5.4. Most interactions occur between side chains and many are hydrophobic in nature. One salt bridge between Syb37 E102 and KDELR R169 is observed, as well as a cation-π interaction between Syb37 R32 and KDELR W120 and some π stacking in edge-to-face mode between several other residues. These can be seen in figure 5.5, which shows a more detailed representation of the interactions of CDR3.

Other Sybodies

Syb37, of the loop family, was the only Sybody which could be successfully crystallised (table 3.5). To investigate the differences and attempt an explanation for why Syb37 is more amenable to crystallography, we can analyse the interactions Syb37 forms and compare this to the sequences of the other sybodies. If we take a look a the sequence alignment of the sybodies in figure 3.6 to determine how unique Syb37 is we can observe that: in CDR1 K31 – R32 are found only in Syb37, other sybodies do not have charged residues here, rather we see hydrophobic and polar residues. CDR2 is less variable, several other sybodies have positive charges in position 53 and W58 is also present in several. When it comes to CD3 sequence comparison between different families is less useful. As the library design differs in CDR3 length and intended binding mode (**Zimmermann et al.**, 2018a).

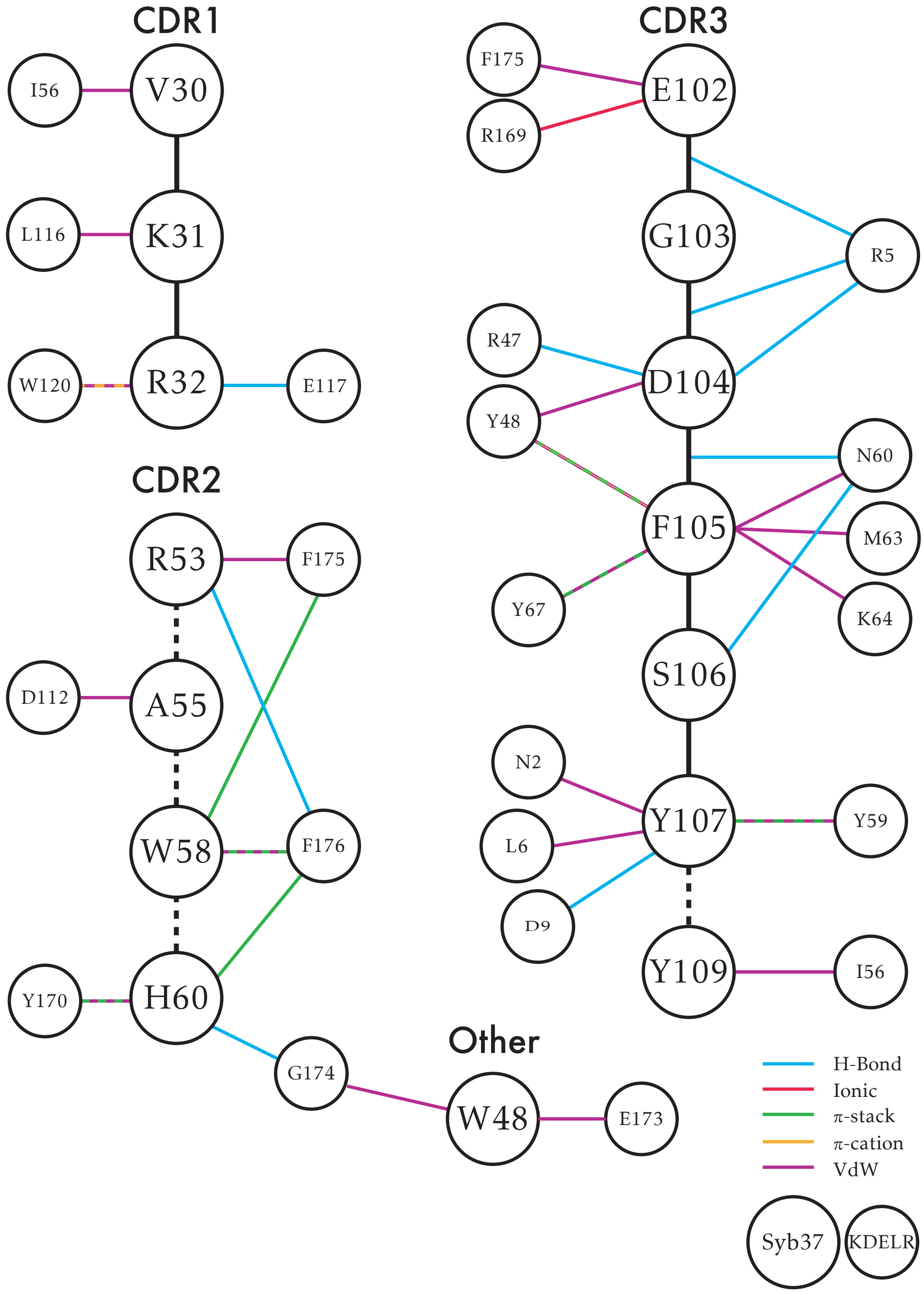


Figure 5.4 Illustration of the interactions between the residues of Syb37 and the KDELR. Side chain to side chain interactions are drawn between circles, interactions between backbone and side chain are drawn between circles and the lines connecting these. Dashed black lines indicate that some residues were skipped in the sequence as they do not contribute to the interaction. Interactions were classified using the RING 2.0 web server (Piovesan et al., 2016).

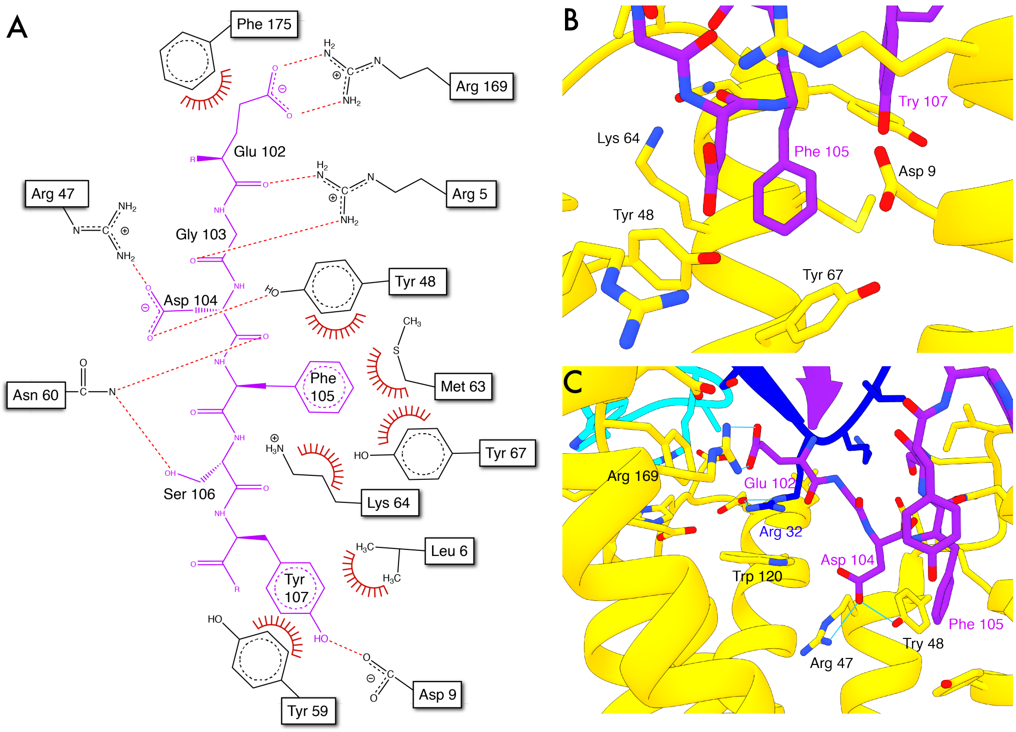


Figure 5.5 **A** A 2D representation of the interactions between the KDELR and Syb37 CDR3. **B** The structure of the interaction around F105 of Syb37 (in purple). **C** The structure of the interaction of another part of CDR3, with some CDR1 seen in the background in dark blue.

Within the medium library, which was designed with loops that enter cavities in mind, we can observe several key differences. Syb37 E102, which forms a salt bridge with KDELR R169 is not present in the others, nor is D104. F105 is only observed in Syb37 and Syb47. Two other Sybodies have aromatic residues at position 107. These observations are confirmed by the phylogenetic tree of the sybody families in 5.6. While the small group of sybodies successfully selected against the KDELR makes analysis difficult, with sybodies generally clustering into the libraries to which they belong, we can see that Syb37 is most closely related to Syb36, followed by the other members of its library.

Comparison of the sequences of Syb36 and Syb37 reveals a series of differences: K31 – R32 in CDR1 become WQ. In CDR2 R53 is S53, but W58 is conserved. H60 is W60. In CDR3 E102 becomes I102, depriving Syb36 of a salt bridge. Residues 104–107, DFSY are SHVW so some aromaticity is found in both loops.

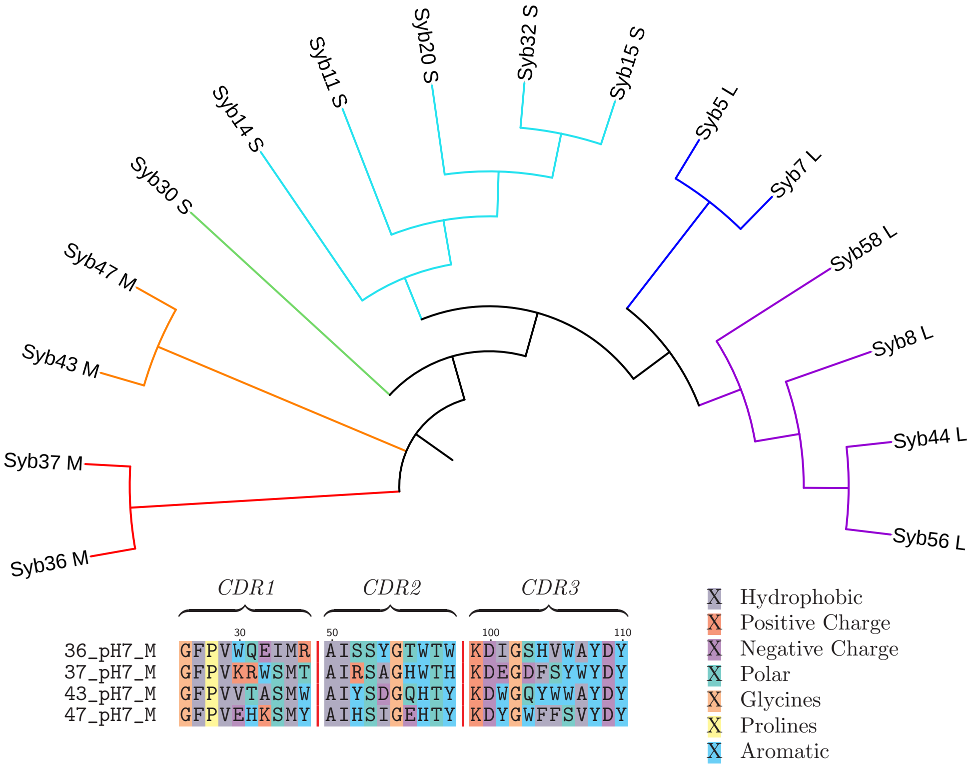


Figure 5.6 Phylogenetic tree of Sybody families. Sybodies cluster together with members of the same library. Sybodies cluster into their three library families, with a total of six smaller subfamilies, indicated by the colours. Below is the sequence alignment of the CDR regions of the medium library.

Given that Syb37 is a member of the medium/loop library, whose binding mode was designed with the CDR3 entering cavities in mind, the scaffold being based on a GPCR bound nanobody (PDB: 3P0G), we can hypothesise that the other members of this library bind in the same mode, presumably in the same polar cavity of the KDELR. It is difficult to come to any conclusions regarding the other two libraries, given their rather different theoretical binding modes, one can only speculate.

5.2.2 TAEKDEL Peptide Binding

The KDELR structure could be solved in a TAEKDEL peptide-bound state at an acidic pH of 6.0. The electron density of the peptide allows unambiguous modelling of the KDEL portion, while density is observed up to the alanine (figure 5.7). The interaction surface between peptide and receptor is 546.3 Å^2, as determined by PISA (Krissinel & Henrick, 2007). The peptide forms a large number of hydrogen bonds and salt bridges with the KDELR, as well as engaging in some hydrophobic interactions. The key residues involved are shown in figure 5.8, many of which appear to be contributed from helix six.

The peptide adopts an orientation perpendicular to the membrane plane, with the lysine making the first substantial interaction to the receptor via a salt bridge to D117 and a cation-πinteraction with W120 (helix 5). The aspartate of the KDEL sequence makes another salt bridge to R169 (helix 6), whereas the glutamate interacts via a third salt bridge to R5, which adopts a different rotamer configuration relative to the apo structure and via a hydrogen bond to W166. The side chain of the C-terminal leucine sits in a hydrophobic cavity formed next to helix 3 and is bracketed on either side by the CH_2 groups of K64 and M63. The carboxy terminus is well coordinated in the binding site, making several interactions to highly conserved residues. In particular two salt bridge interactions made to R47 (helix 3) and R159 (helix 6), anchor the C-terminus of the peptide in the binding site.

The extensive mutational analysis by Townsley et al. (1993) and Scheel & Pelham (1998) identified many KDELR mutants that disrupted the binding to a KDEL peptide sequence. We observe many of these residues directly in the binding pocket, including (R5, R47, K64, R159, Y162, N165 and R169), other residues they identify as important for binding occur near the binding site (D9 and D50) while some appear to be very distant from the binding site (I32 and D91).

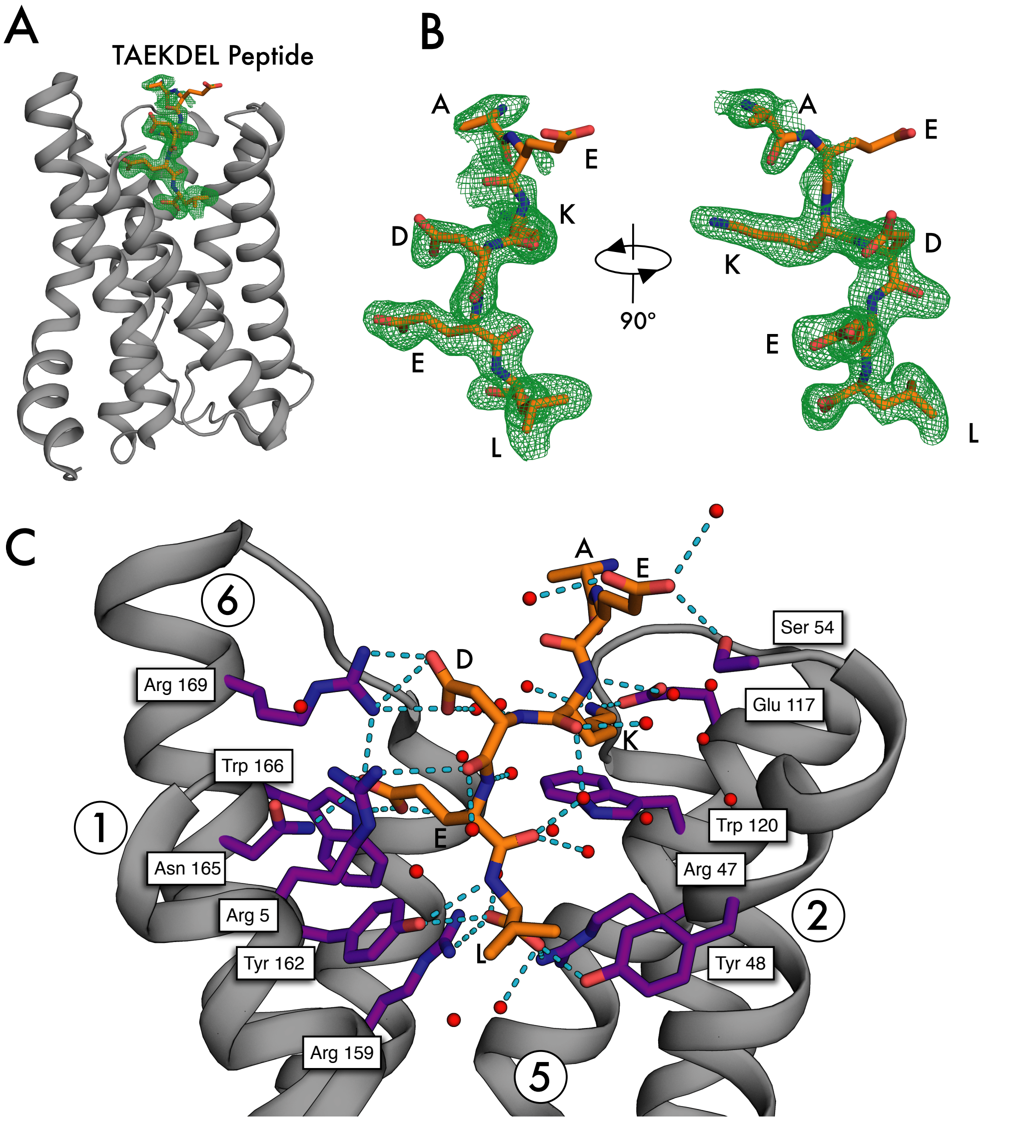


Figure 5.7 **A** Polder OMIT map of the TAEKDEL shown in green mesh, contoured at 3σ in the context of the KDELR. **B** View of the same density without the KDELR, the density for the threonine of the TAEKDEL peptide is not observed. **C** Binding site of the KDELR with some key residues shown in purple, and waters shown in red. Circled numbers indicate the TM helix number.

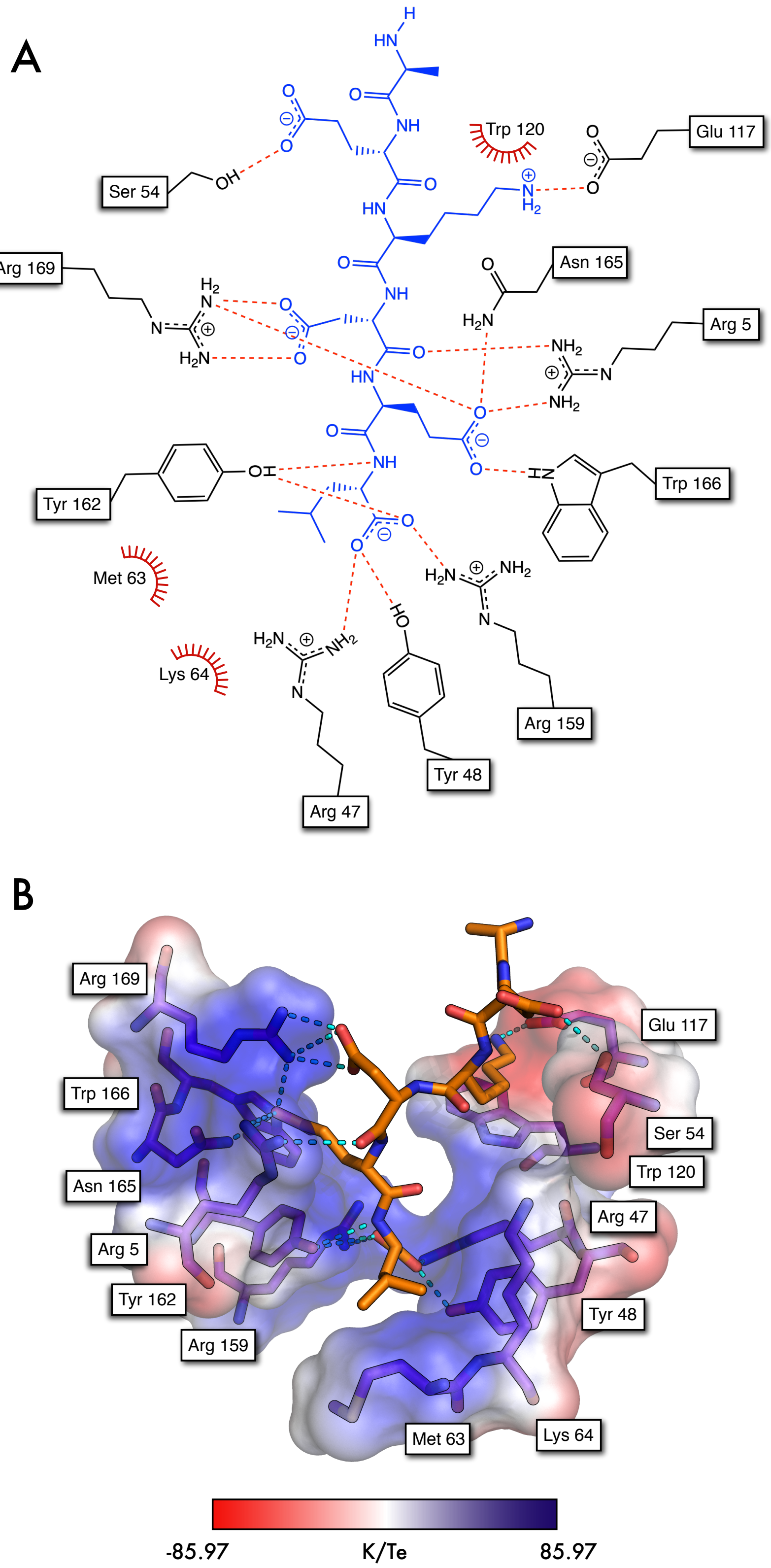


Figure 5.8 **A** 2D Representation of interactions between TAEKDEL peptide and the KDELR. **B** Residues shown as they are in the KDELR, with the electrostatic surface of the binding pocket being shown.

5.2.3 TAEHDEL Peptide Binding

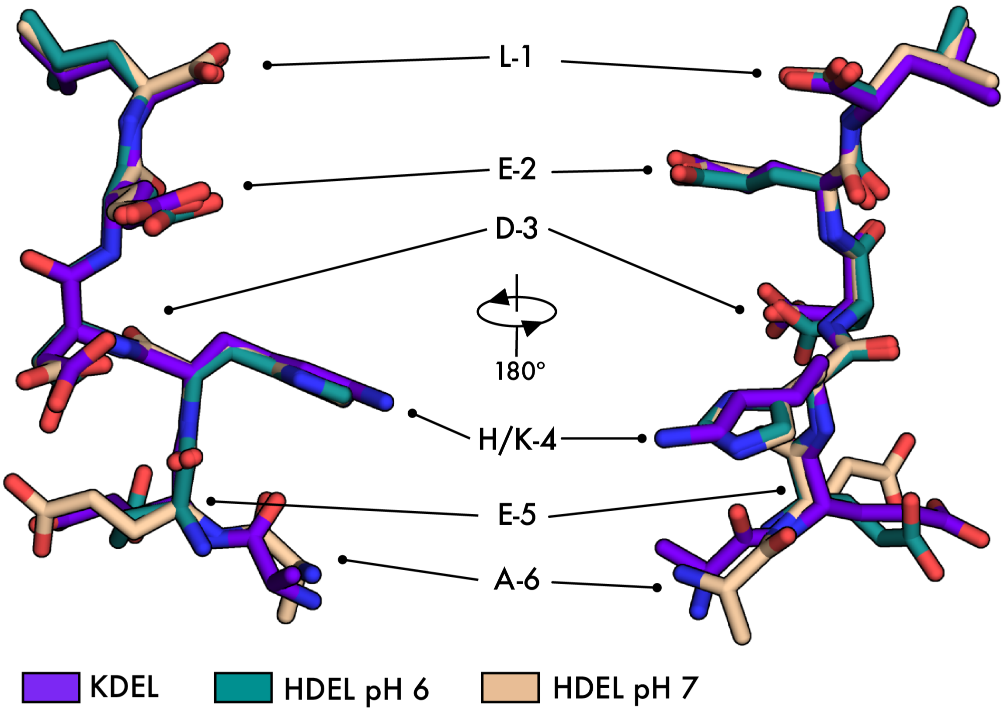


Figure 5.9 Superposition of KDEL, HDEL pH 6 and HDEL pH 7 peptides, the pose of the peptides is virtually identical, after the -4 position more variability occurs as this region is outside the binding pocket.

TAEHDEL was observed bound to KDELR in an acidic and a neutral environment. In both cases the binding pose is identical to that of the TAEKDEL peptide shown in figure 5.7. Figure 5.9 shows a superposition of all three peptides. The RMSD between pH 6.0 TAEHDEL and the TAEKDEL structure is 0.128 Å over 198 Cα, the difference between the pH 7.0 TAEHDEL and the TAEKDEL structure is 0.157 Å over 197 Cα. The RMSD between the two TAEHDEL structures is 0.198 Å over 203 Cα. While the RMSD between the pH 6.0 and apo, and pH 7.0 and apo is 0.601 Å over 152 Cα and 0.673 Å over 155 Cα respectively.

5.2.4 HDEL vs KDEL Binding Mode

The key difference between the binding of the TAEHDEL as compared to the TAEKDEL peptide, is the exchange of a histidine for a lysine residue at the -4 position. This position of the peptide motif makes interactions with two residues of the receptor that are seen in figure 5.10. A salt bridge is formed

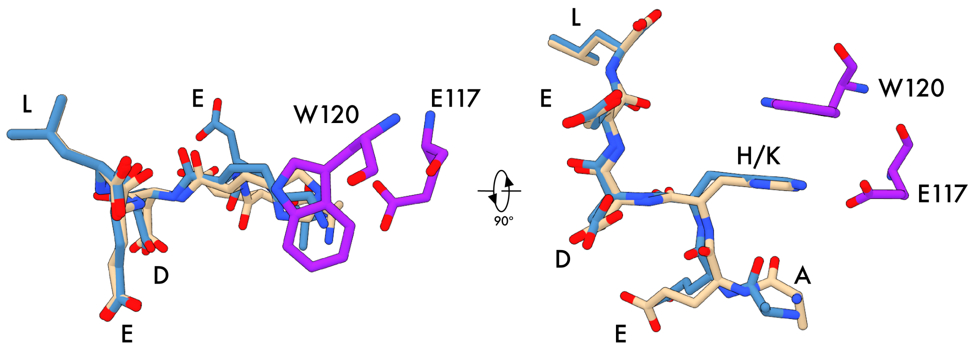


Figure 5.10 TAEHDEL and TAEKDEL peptides overlaid, with the W120 and E117 residues of the KDELR shown.

between E117 and the positive charges present on lysine and histidine. Lysine also forms a cation-π interaction with W120, which histidine engages in a mixed π-π, cation-π interaction. It is these two residues that must form the basis of selectivity between HDEL and KDEL motifs, as no other residues are close enough to make meaningful interactions with the -4 position of the KDEL motif in this pose. The nature of these interactions will be explored further in section 6.5. The interaction energies, especially that of lysine with W120 compared to histidine (neutral or protonated) cannot be trivially compared and will require calculation. Protonated histidine has been reported to have significant π-π stacking energies, though cation-π interactions from lysine residues can make appreciable energetic contributions, so comparison of their contributions to peptide binding will likely be very context specific (Liao et al., 2013; Gallivan & Dougherty, 1999, 2000).

5.2.5 pH 7 vs pH 6 Binding Mode

Globally the difference between the pH 6 bound and pH 7 peptide-bound structures is minimal, with the receptor adopting the same conformation in each of these structures. One key difference however, the importance of which will become clear momentarily, is the length of a hydrogen bond between Y158 and E127. In both pH 6 peptide structures the length is observed to be ~2.5 Å, while in the pH 7 structure this appears to be a little longer, with a length of ~2.7 Å -

though differences of this size, given resolution of the structures, are plausibly within error, given that the expected maximal error in the coordinates of the pH 7 structure is 0.352 Å, and 0.164 Å for the TAEKDEL structure (**Vaguine et al., 1999**).

5.2.6 Syb37 vs Peptide Binding

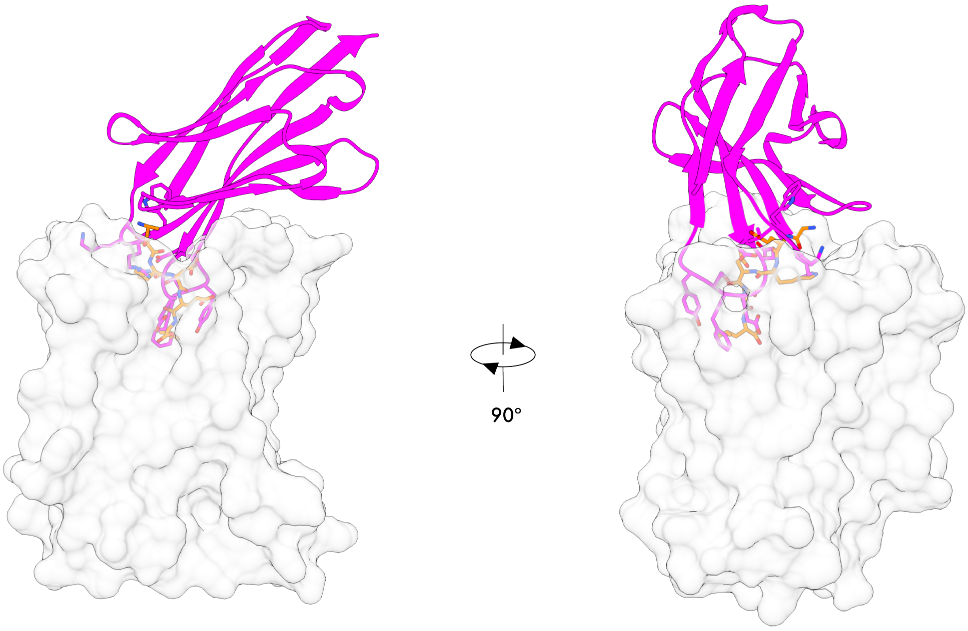


Figure 5.11 Overlay of Syb37 onto the the KDELR TAEKDEL structure. CDR3 of Syb37 (pink) and the TAEKDEL peptide (orange) make use of the same pocket for binding.

Syb37 and the KDEL peptides bind in the same polar cavity of the KDELR (figure 5.11). The mode of interactions is largely different, with Syb37 making more hydrophobic interactions. However, given the constrained size of the pocket, some similarities do arise: Syb37 D104 interacts with R47, similar to the carboxy terminus of the Leucine and F105 makes hydrophobic contacts with M63 and K64, just as the side chain of the Leucine. Syb37 R32 forms a cation-π interaction with W120, as do the H/K of the peptide. Finally, R169 is engaged by Syb37's E102, while the peptide's aspartate also forms a salt bridge here.

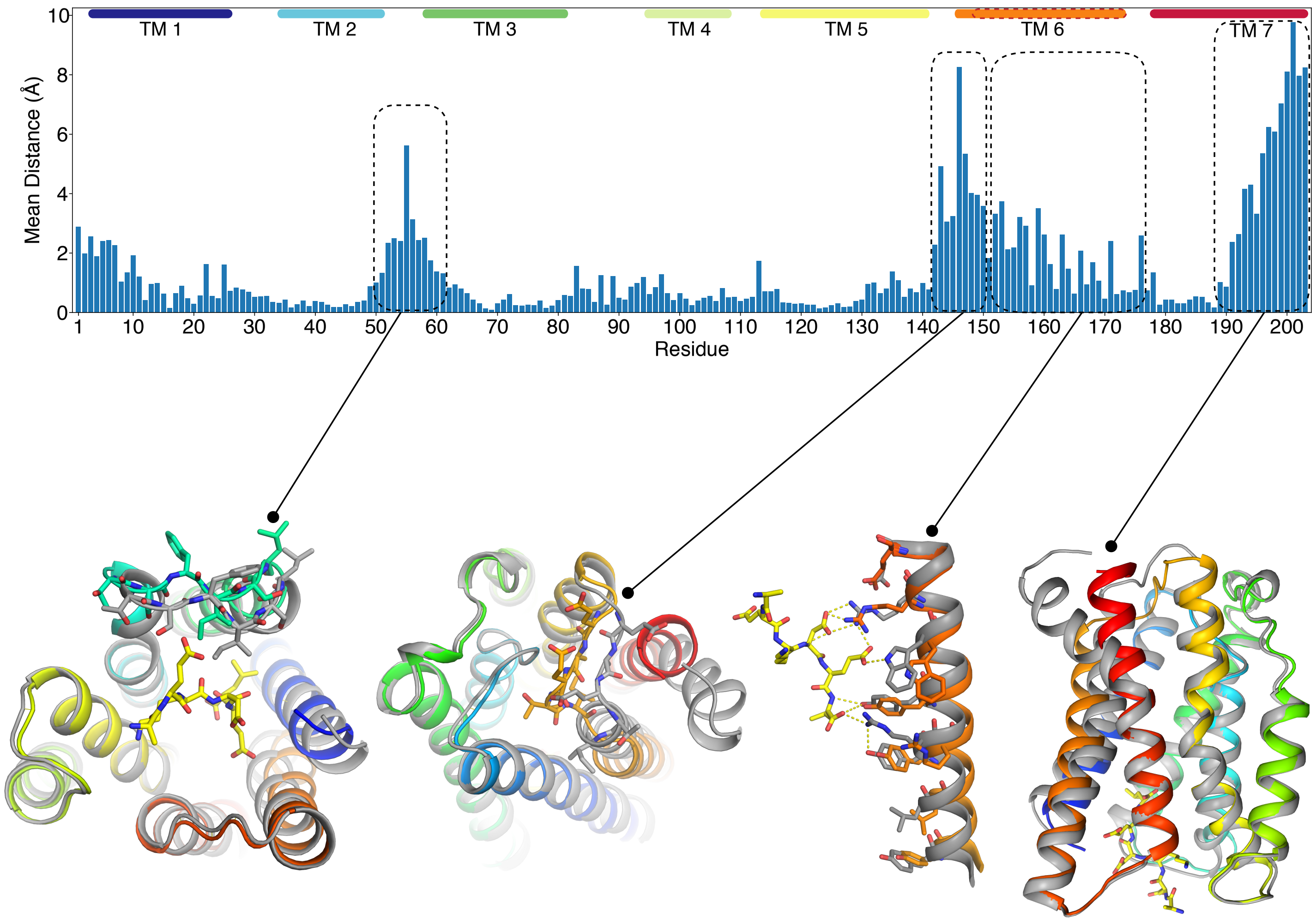


Figure 5.12 Plot showing the distance between the geometric centres of the residues of the apo and peptide-bound KDELR, indicating which parts of the KDELR move the most upon peptide binding. The most mobile areas are shown on the structures below. The peptide-bound structure is shown in grey, while the apo structure is seen in a rainbow colouring. The positions of the helices are indicated by the coloured regions above the plot, the dashed red line indicates the size of helix 6 in the apo state, which is shorter than in the bound state. Of the residues with significant distances between bound and apo states, only the very C-terminal lysines in the peptide-bound structure engage in crystal packing and can be seen in the inset of figure 4.8, these lysines are not observed in the apo structure, so they are not shown in the plot above.

5.3 Bound vs Apo Structure

We have seen that the peptide-bound structures are broadly similar to each other, but they demonstrate key differences to the apo structure. In figure 5.12 the regions of largest difference are highlighted. They cluster into two general areas on the receptor, the cytosolic face and the residues in helix six, engaging the bound peptide.

5.3.1 Peptide Binding Pocket

The residues engaged in peptide binding were explored in figure 5.8. Figure 5.13 shows a structural overlay of these same residues from the apo and TAEKDEL structures. Many of the residues in this region move, though some travel larger distances upon peptide binding. In particular residues W166 (2.8 Å), Y158 (1.2 Å) and R159 (4.8 Å) on helix 6, which move in to meet the peptide, and R5 (4.7 Å) and D9 (1.6 Å) on helix 1, which move out to make space, have the most pronounced conformational changes. The loop between helix 2 and 3 also appears to move in slightly towards the peptide in the bounds state though it does not appear to have any significant interactions with the peptide.

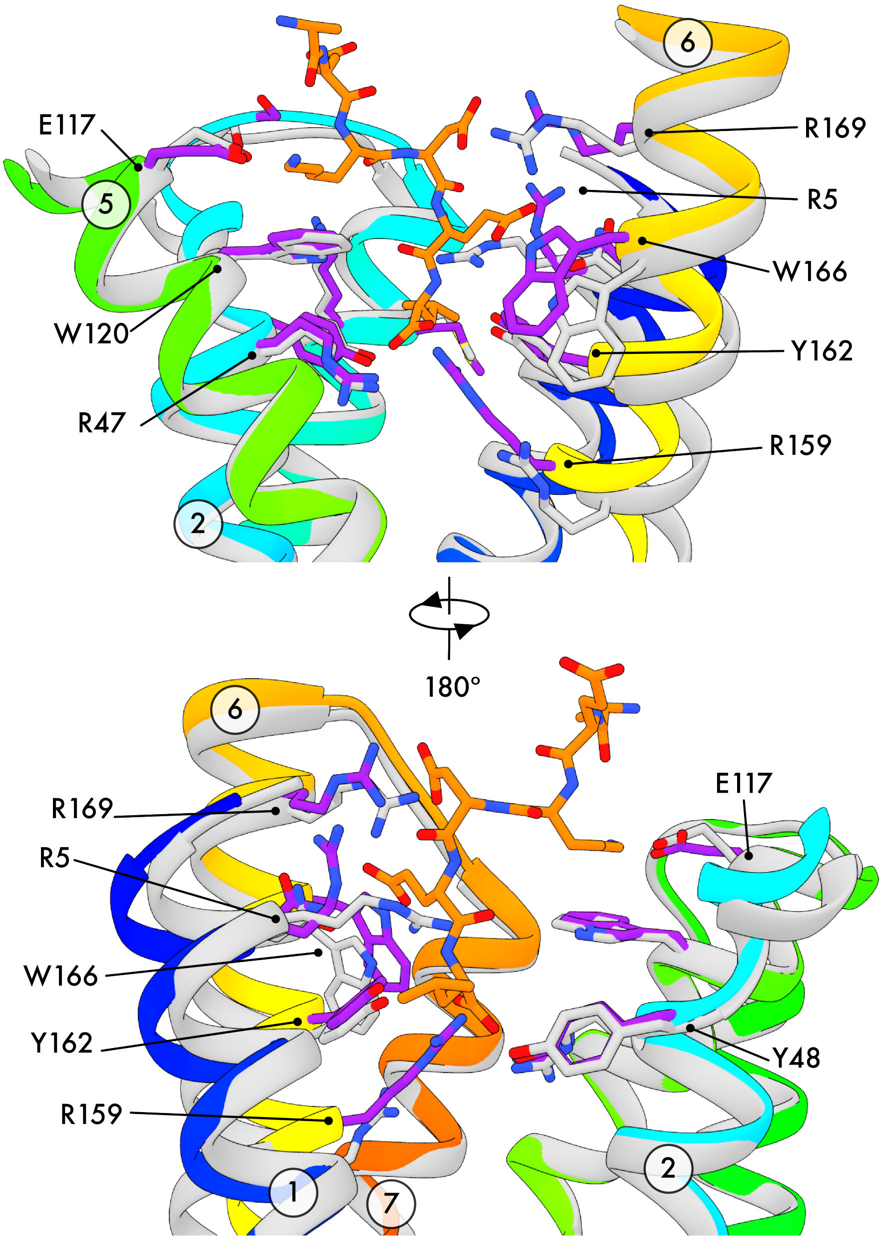


Figure 5.13 The TAEKDEL bound KDELR structure (rainbow, purple side chains) and apo structure (grey) showing the movement of residues upon peptide binding.

5.3.2 Cytosolic Face

On the cytosolic side of the KDELR the chief differences in apo and peptide-bound structures lies in the second THB. Most noticeably helix 6 gains a turn in the bound state, thus also altering the loop connecting it and helix 5. The most pronounced difference is the kinking out of helix 7 by 14 Å, with Y191 being the first residue with a marked shift in position, see figure 5.14.

Potential COPI & COPII Binding Sites

If we observe the difference in electrostatic potential between the bound and apo state on the cytosolic side, we notice some pronounced differences. It must be noted here that the model for the apo structure ends at residue 203, after which no interpretable density is observed, while the peptide-bound structure is modelled until residue 207. Thus K204, 206, 207 are not observed in the apo form. In the apo form, we see a band of negative charge running across the cytosolic face of the KDELR. The charge is contributed by several invariant acidic residues in the mammalian KDELRs, including D87, E143, E145 and the C-terminus of helix 7 in the apo model. Mutagenesis studies implicated several residues in the cytoplasmic portion of human KDELR that result in complete retention of the receptor in the ER (Townsley et al., 1993), i.e. unable to exit the ER in COPII vesicles. Interestingly, all of the mutated residues function to support the structural integrity of this electrostatic feature, suggesting this may form part of a di-acidic COPII recognition motif (Barlowe, 2003).

Upon binding of KDEL peptide we observe a movement of helix 6. This has a pronounced effect at the cytoplasmic end of the receptor, causing helix 7 to move away from helix 5, creating a new cavity on the cytoplasmic side of the membrane. The new position of helix 7 results in the repositioning of a strictly conserved acidic residue, D193 (in helix 7). A D193N mutant was previously shown to bind KDEL peptide *in vitro*, yet remains trapped in the Golgi *in vivo* in the presence of an excess of KDEL ligand (Townsley et al., 1993). This suggests that the conformational change in helix 7 and charge at this position is important for ER retrieval. Movement of helix 6 and 7 radically alters the electrostatic surface at the cytoplasmic end of the receptor, resulting in the negatively charged band being split and truncated. This movement also results in the exposure (or structural ordering) of previously buried lysine residues, K201, 204, 206 and 207, at the C-terminus of helix 7 and the formation of a cytoplasmic cavity. The partially exposed lysine cluster is reminiscent of the C-terminal KK*xx* and K*x*K*xx* di-lysine motifs, which are important in COPI-dependent Golgi to ER transport (**Letourneur et al.**, 1994; **Jackson et al.**, 2012). These lysine residues are not resolved in the apo structure.

To assess the structural plausibility of these motifs, one can compare the location of the postulated retrieval motifs on the KDELR with either known structures of recognition motifs or of the architecture of the COP in question. In figure 5.15 we see the structure of the COPI coat from subtomogram averaging, the α-COP and β′-COP β-propellers are directly facing the membrane and are in an ideal position to contact the lysine motif on the KDELR (**Bykov et al.**, 2017). For COPII motifs the structural recognition seems less obvious for the acidic residues proposed above, as we saw in figure 1.2 Sec23-Sec24 recognises the syntaxin-5 peptide at some distance from the proposed location of the membrane. The structure of cystic fibrosis transmembrane conductance regulator (CFTR) which has a known DAD COPII motif presents these residues a considerable distance from the membrane (**Wang et al.**, 2004; **Liu et al.**, 2017). In the KDELR the proposed residues sit fairly close to the membrane, so how these would directly engage COPII machinery is not clear from the current structural information. It may be further noted that neither the acidic residues or the lysine residues are strictly conserved in KDELR homologues.

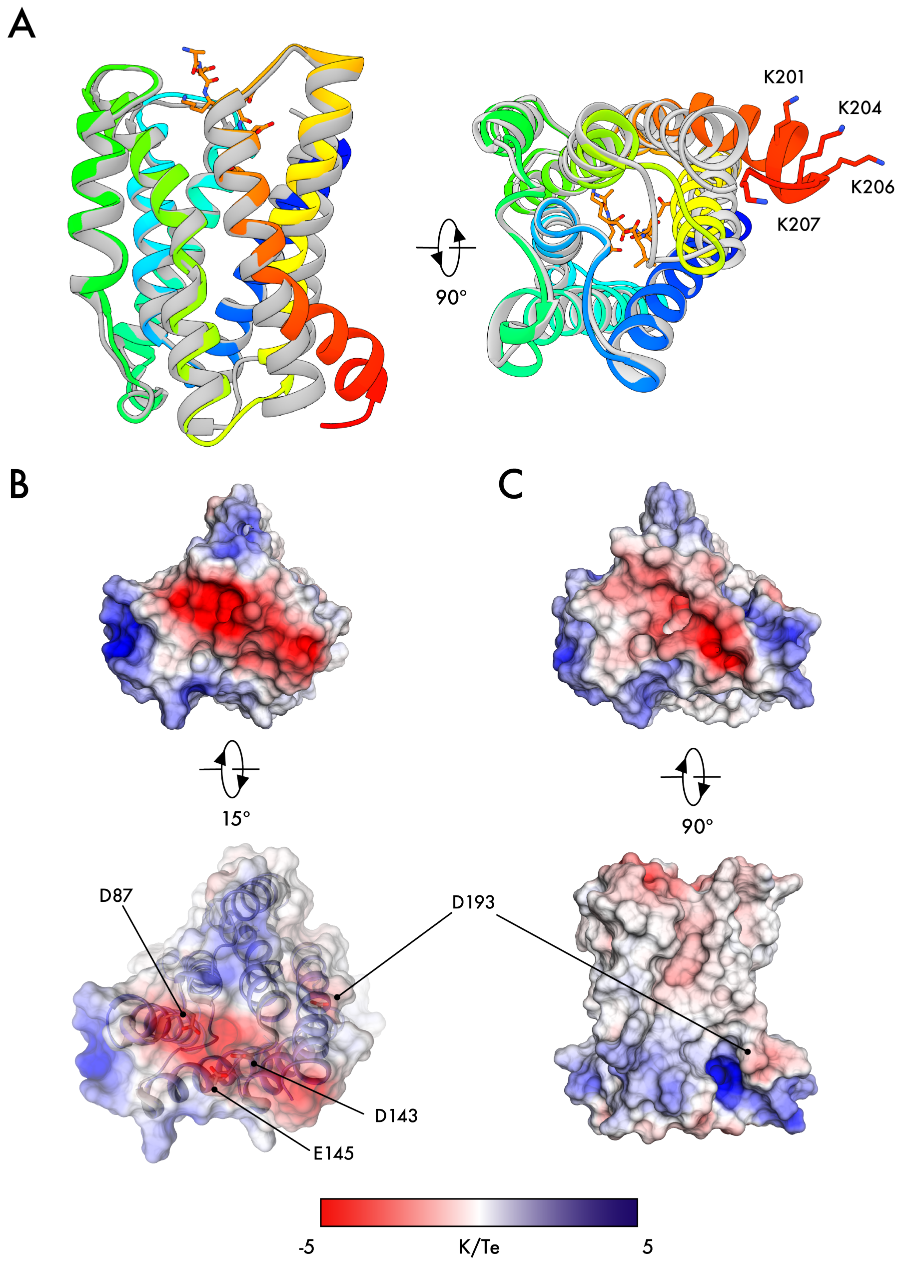


Figure 5.14 **A** Overlay of bound (rainbow) and apo (grey) KDELR structures. The movement of helix 7 and the winding of helix 6 can be seen clearly, and the terminal K residues are shown in stick representation. **B** Electrostatic surface representation of the apo structure, below showing the key residues forming the negatively charged band. **C** Electrostatic surface representation of the TAEKDEL bound structure, below indicating the position of residue D193. Electrostatics calculated with APBS (Jurrus et al., 2018).

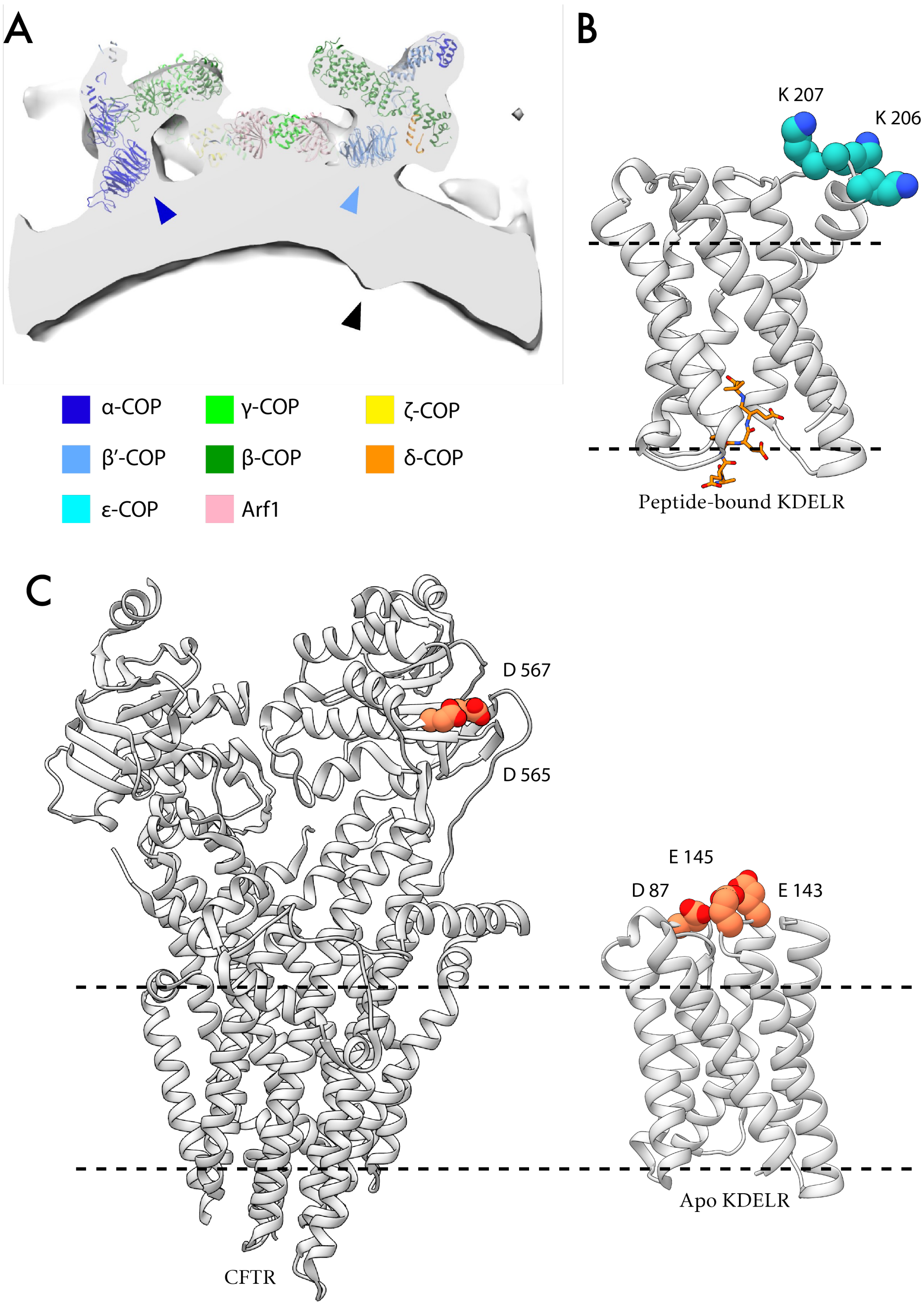


Figure 5.15 **A** The COPI coat from subtomogram averaging shows the location of α-COP and β′-COP relative to the membrane, with the black arrow indicating potential cargo density, adapted from Bykov et al. (2017). **B** The peptide-bound structure of KDELR highlighting the location of the putative COPI lysine retrieval motif. **C** The structure of CFTR (PDB: 5UAK (Liu et al., 2017)) with its di-acidic COPII motif, compared to the acidic residues that could be part of a putative COPII motif on the apo KDELR. Dashed lines indicate the membrane bilayer.

5.4 pH-Dependent Binding Mechanism

From an analysis of the residues forming the binding pocket of the peptide, no obvious mechanism for pH-dependence was observed. None of the residues present had any titratable protons in the range the KDELR would display sensitivity to pH (pH 6–7). However, the short hydrogen bond between E127 and Y158 noticed during refinement, has a flanking histidine residue, which would be titratable, having a theoretical pK_a around pH 6.

5.4.1 An Unusual Hydrogen Bond

In the apo form the distance between the oxygens of E127 and Y158 is 3.2 Å, while in the peptide-bound structure this distance is reduced to 2.5 Å. Y158 sits on helix 6 and appears to follow the movement of this helix towards the bound peptide, possibly driven by the large movement of R159 to engage with the C-terminal carboxy group of the peptide. Hydrogen bonds with distances approaching 2.5 Å have much higher $\Delta G_{\text{formation}}$ than longer ones, -25–7 kcal mol^{-1} vs -3–1 kcal mol^{-1} for those longer than 2.7 Å (Harris & Mildvan, 1999; Cleland & Kreevoy, 1994; Jeffrey, 1997).

In weaker hydrogen bonds, the hydrogen is attached to one of the oxygens in the acceptor-donor pair covalently. If the pK_a of the two oxygens is similar the hydrogen may be associated with either oxygen, forming a symmetrical hydrogen bond. If the distance between the oxygens decreases, the energy wells in the potential function of the hydrogen position merge from two separated wells into two wells with a low barrier which the hydrogen may overcome, so called low-barrier hydrogen bonds (Cleland & Kreevoy, 1994). Hydrogen bonds can be theoretically classed by the shapes of their potential wells, which are shown in figure 5.17. It must be noted that the presence of low-barrier hydrogen bonds in proteins is contentious and that as the crystal structure of the KDELR does not allow the determination of the hydrogen positions, the hydrogen bond can only be characterised by quantum mechanic (QM) calculations, neutron diffraction (which is implausible with LCP crystals) or NMR experiments (Yamaguchi et al., 2009; Agback & Agback, 2018). For this reason the hydrogen bond here shall be termed a *short hydrogen bond*.

Formation of this short hydrogen bond could stabilise the bound position of helix 6, locking the peptide in place through its interaction with R159. Y158 sits close to the conserved H12 on helix 1, forming a face-to-face aromatic interaction. Two water molecules sit at the bottom of the peptide-binding site, coordinating the peptide carboxyl group to both Y158 and H12, which are further stabilised through a hydrogen bond to D9 on helix 1. Given its solvent-accessible position near the peptide-binding site, it is possible that protonation of H12 facilitates the formation of the short hydrogen bond stabilising the position of helix 6, thus acting as the pH sensor within the receptor.

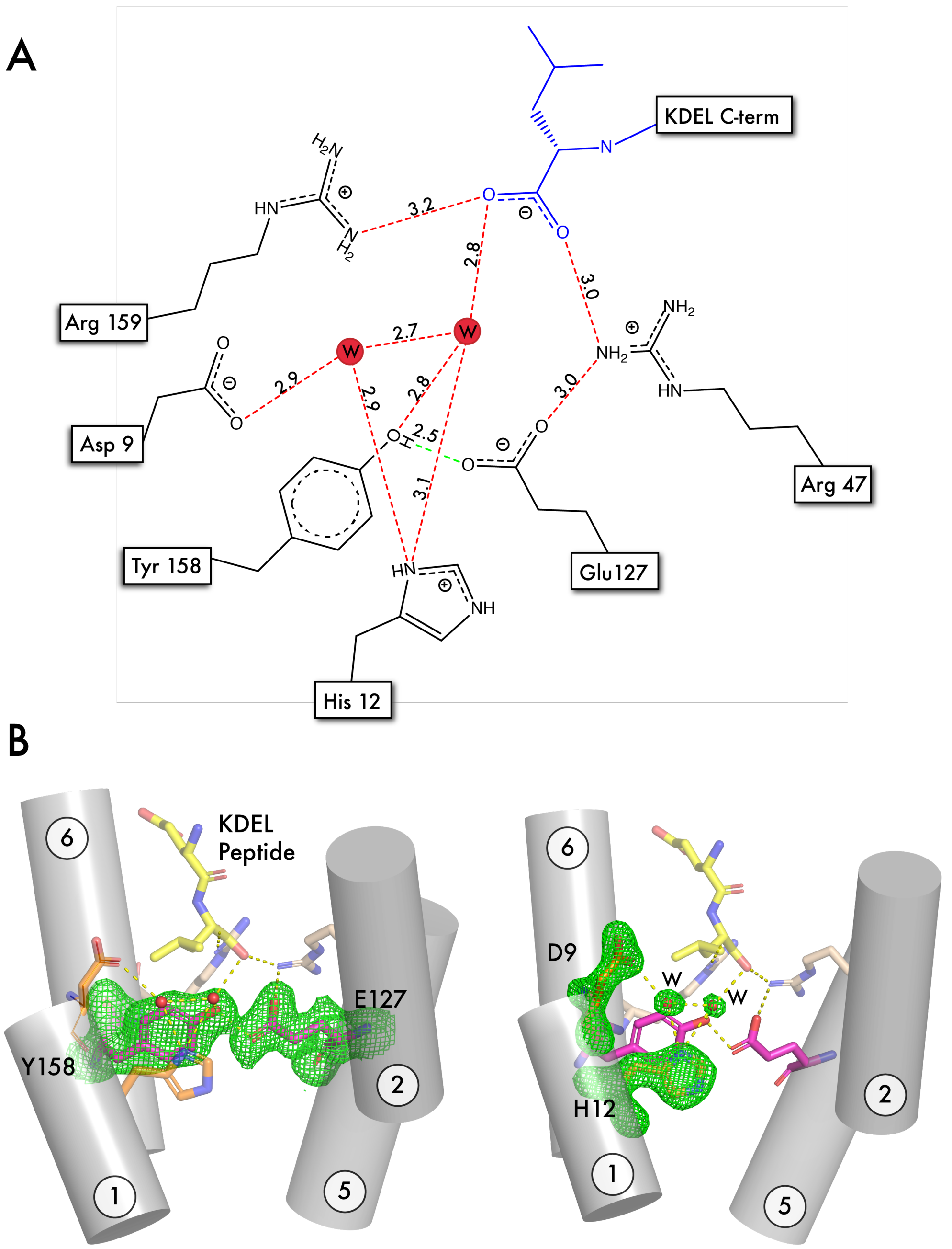


Figure 5.16 **A** Schematic of the residues involved with the hydrogen bond between E127 and Y158 in the TAEKDEL bound structure. **B** On the left, the electron density from a Polder OMIT map is shown (green mesh), contoured at 3σ, around E127 and Y158. Two water molecules are shown (red spheres), connecting Y158 to H12, suggesting a possible mechanism for pH sensing in the ER and Golgi. The second water molecule is observed forming a hydrogen bond to D9, mutation of which results in a receptor that is severely reduced in KDEL peptide binding (Townsley et al., 1993). Hydrogen bonds shown as dashed lines (yellow). R47 and R159 (wheat) are shown coordinating the carboxy terminus of the peptide. To the right, the equivalent view showing the Polder OMIT map (green mesh), contoured at 5σ, around H12 and D9 and the two water molecules (W) linking these residues to Y158 and the carboxy terminus of the TAEKDEL peptide.

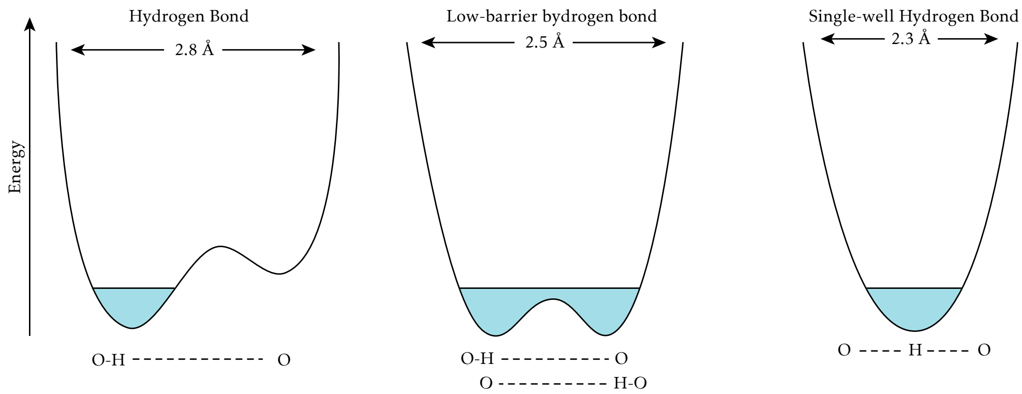


Figure 5.17 Potential functions of a hydrogen bond. Most hydrogen bonds in proteins are asymmetric and the hydrogen will be covalently bound to the donor atom. If the pK_a of the donor-acceptor pair is equal, a symmetric hydrogen bond may form. In some instances when the distances are small the potential barrier between the donor and acceptor shrinks and a low-barrier hydrogen bond may form, where the hydrogen may move between both oxygens. An even shorter distance leads to a single-well hydrogen bond with very short distances, these are not thought to appear in proteins. The horizontal lines represent the lowest energy levels for hydrogen (**Cleland & Kreevoy**, 1994; **Agback & Agback**, 2018).

5.5 Comparison to Transporters and Receptors

The KDELR is a member of a wider family of proteins known as the PQ-loop family, named after a conserved PQ di-peptide (Saudek, 2012). This family has been proposed to belong to the TOG superfamily (Yee et al., 2013). The role of the PQ motif is not clear, currently only SWEET protein structures are available on the PDB. Some have suggested PQ-loop proteins are involved in small molecule transport, as cystinosin, mitochondrial pyruvate carrier, and SWEET proteins all perform this function (Jézégou et al., 2012). In semiSWEET proteins, the dimeric ancestral bacterial precursor (two triple helical bundles) to the eukaryotic SWEET family, the PQ-loop acts as a molecular hinge, important in the transport cycle (Lee et al., 2015). LAAT-1 a lysosomal lysine arginine transporter requires the conserved PQ for transport (Liu et al., 2012).

With this information in mind the KDELR was analysed in relation to structurally similar proteins, as well as GPCRs, as the KDELR has in fact been proposed to act as a GPCR (Giannotta et al., 2015).

5.5.1 PQ-loop

The hinge motion of the PQ-loop observed in semiSWEETs is not seen in the KDELR, between the apo and bound structures there is little movement in this region. The PQ motif sits on helix 5, just after E127, which is involved in the short hydrogen bond. The region around the PQ-loop, especially between E127 and L134 is highly conserved (figure 5.18). The residues between E127 and P133 also show some unwinding of the helix (Kabsch & Sander, 1983). Given the high degree of conservation of the PQ-loop and helix 5 in general, it must be important for the function of the KDELR. It does not appear to act as a molecular hinge, and given that the KDELR does not transport across a membrane, this would not be required. In the mutational analysis of Townsley et al. (1993), a Q134A mutation did not cause any distributional effects in their retrieval assays, but no binding experiment was performed. Cystinosin was found to require its

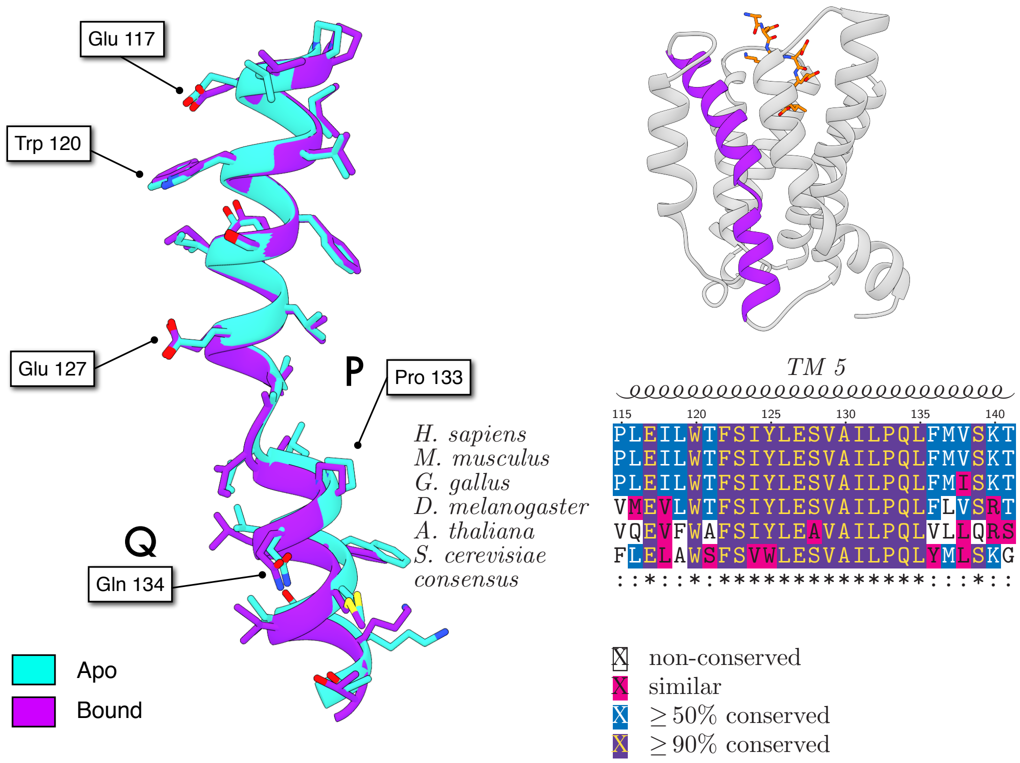


Figure 5.18 Helix 5 of the bound and apo KDELR, showing some of the conserved residues and the PQ di-peptide. On the right hand side an alignment of KDELRs shows the high conservation of this helix.

PQ motif for lysosomal targeting (Cherqui et al., 2001). This leaves the the role of the PQ-loop in the KDELR an open question.

One of the few proteins related to the KDELR for which functional data exists is cystinosin. Cystinosin is thought to be a proton coupled cystine symporter, which removes cystine from lysosomes in a pH-dependent manner (Kalatzis et al., 2001). Cystinosin's PQ-loop may be involved in lysosomal targeting and it appears that when cystinosin is absent LAMP2A is mis-localised, indicating a potential trafficking role for cystinosin (Cherqui et al., 2001; Saudek, 2017; Zhang et al., 2017). As mentioned above, the role of the PQ-loop in the KDELR is unclear, however, as more family members are investigated this question may be answered, making cystinosin an interesting target for structural studies.

5.5.2 Transporters: SWEETs and PnuC

The KDELR shows some structural similarity to other members of the PQ-loop family, the SWEETs. SWEET proteins are sugar transporters that catalyse facilitated diffusion of sugars along their concentration gradients (Jia et al., 2017). Eukaryotic SWEETs have seven trans-membrane helices, arranged in the same topology (3+1+3) as the KDELR. Prokaryotic semiSWEETs are thought to be an ancestral form of the SWEETs and consist of two dimers of three transmembrane helices. It is thought that SWEETs derived from semiSWEETs in a gene duplication event, with a linker helix being inserted to maintain the correct topology (Feng & Frommer, 2015; Jaehme et al., 2015). When aligned with *Os* SWEET2b, *Gg* KDELR2 shares about 25.6% identity, with many gaps in the alignment. Aligning *Gg* KDELR2 with two copies of *E. coli* semiSWEET shows an identity of 21.6%. Structural alignments of KDELR onto SWEET2A (PDB:5CTG (Tao et al., 2015)) show a high RMSD (6.2 Å for both peptide-bound and apo KDELR) for the whole structure but this improves if alignments are made only over the THB, decreasing to 2.56 Å for THB2 (figure 5.19). When the apo KDELR is aligned to a semiSWEET structure (PDB: 5UHS (Latorraca et al., 2017)), the RMSD is 2.47 Å over 80 Cα. Interestingly, this structure of the semiSWEET was solved in the presence of glucose, which appears to partially overlap with a cavity in the KDELR.

Others have noticed that the vitamin B3 transporter, PnuC, shares a similar architecture to SWEETS (Jaehme et al., 2015). PnuC has a different topology to SWEETS, with a sequential order of helices. However, when the structures are overlaid, a six helical core, consisting of two THBs emerges. It has been proposed that this may have evolved through a 3D domain swap event, though it is currently not possible to prove the con- or divergence of transporters on this THB dimer architecture (Jaehme et al., 2015, 2016; Feng & Frommer, 2016). Sequence alignment with *Neisseria mucosa* PnuC and the KDELR shows an identity of about 21.9%. PnuCs six-helical core can be overlaid with the KDELR core (RMSD 5.27 Å), and the vitamin bound to the PnuC structure also overlays

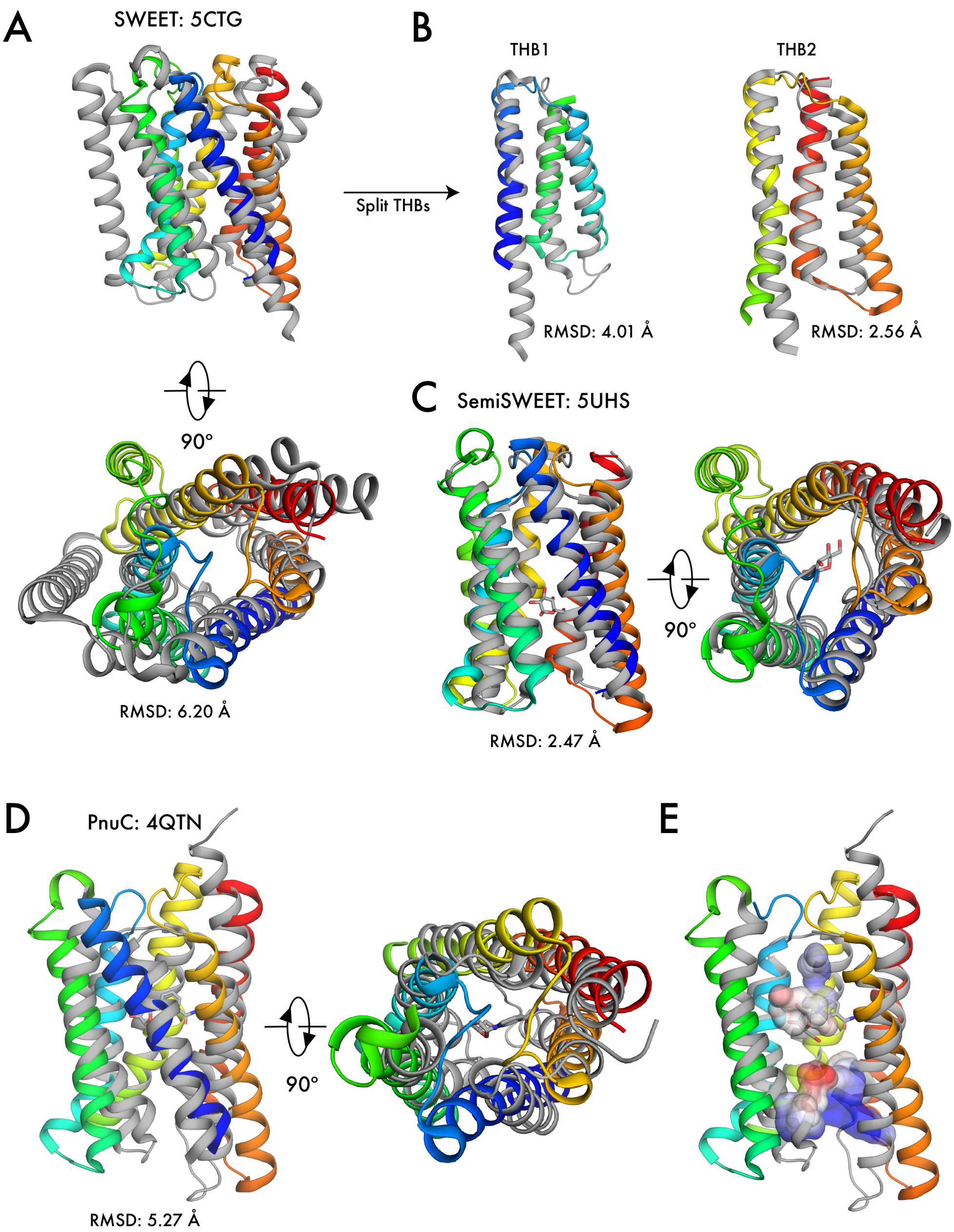


Figure 5.19 **A** Alignment of apo KDELR with SWEET2b (PDB: 5CTG (Tao et al., 2015)), gives an RMSD of 6.20 Å over 167 Cα. **B** Aligning the THB of KDELR and SWEET2b shows a much lower RMSD of 4.01 Å over 72 and 2.56 Å over 80 Cα for THB1 and 2 respectively. **C** Alignment of apo KDELR with semiSWEET (PDB:5UHS (Latorraca et al., 2017)), yielding an RMSD of 2.47 Å over 152 Cα. **D** Alignment of apo KDELR with PnuC (PDB:4QTN (Jaehme et al., 2014)), using only the six helical core of both proteins, with an RMSD of 5.27 Å over 107 Cα. **E** The same alignment as in **D**, showing cavities of the KDELR which overlap with the vitamin B3 in the PnuC structure.

with a cavity of the KDELR, which appears to extend to the cytosolic side in the apo KDELR structure.

5.5.3 Receptors

The KDELR has been reported to activate three different G-proteins, $G\alpha_q$, $G\alpha_s$ and monomeric $G\alpha_o$ (Giannotta et al., 2012; Cancino et al., 2014; Solis et al., 2017). Sequence similarity of the KDELR to GPCRs is extremely low (~ 10%), and the KDELR is generally much smaller than GPCRs, which are around 40 kDa (with large variability). It was anticipated that the KDELR would share a seven transmembrane architecture with GPCRs, though no homology was predicted (Lewis & Pelham, 1992a).

The topology of the KDELR is different to that of GPCRs, which do not have the 1-3-2 THB architecture seen in the KDELR. Structural alignments between the KDELR and GPCRs, using only Cα atoms, reveal little similarity beyond the seven helices. RMSDs range from 10 Å to well over 16 Å. In figure 5.20 **A**, an overlay of peptide-bound KDELR with frizzled 4 receptor, a class F GPCR, (6BD4 (Yang et al., 2018)) is shown (10.2 Å over 196 Cα), one of the structures with the lowest RMSDs measured. While some of the structures overlap in the core of the proteins, which are embedded in the membrane, the two proteins differ vastly at the solvent exposed sides. If we try to align peptide-bound KDELR onto a G_s bound GPCR, (3SN6, β2 adrenergic receptor (Rasmussen et al., 2011), RMSD of 13.7 Å over 198 Cα) it is not obvious where a G-protein would bind on the KDELR, perhaps in the cavity of the second THB formed by the kinking of helix 7?

Several GPCRs recognise peptide substrates, such as the glucagon (a peptide hormone) receptors and the neurotensin (a neuropeptide) receptors (Wu et al., 2017). Figure 5.21 shows the locations and orientations of peptides bound to the neurotensin receptor 1 (NTS1) receptor, a Class A GPCR bound to a neurotensin (NT) truncation NT_{8-13}, and glucagon-like peptide 1 (GLP-1), a class B GPCR bound to peptide 5, next to the KDELR bound to a KDEL peptide (White et al., 2012; Jazayeri et al., 2017). All three peptides adopt a perpendicular orientation relative to the membrane plane, NT and KDEL both enter their binding

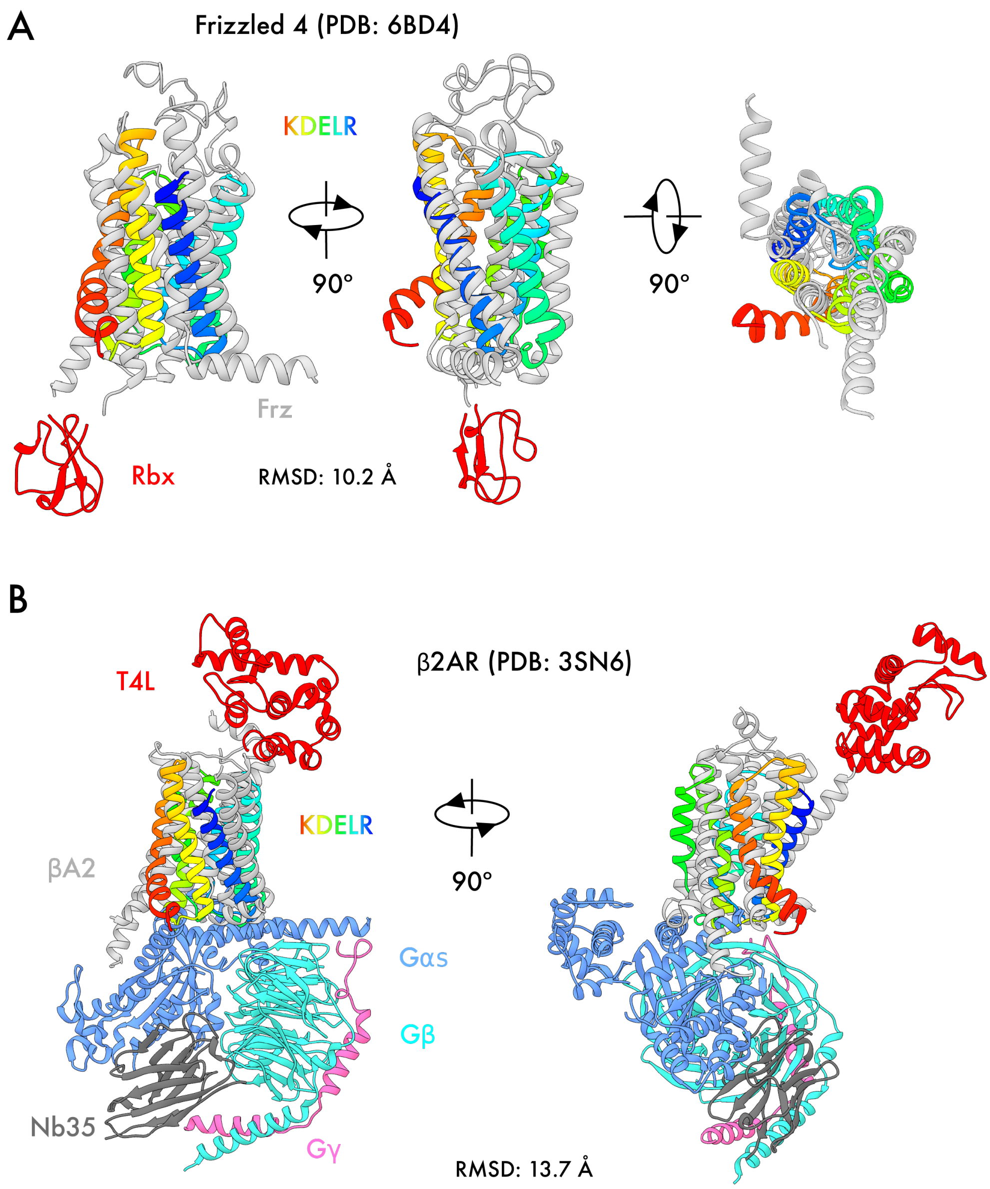


Figure 5.20 **A** Structural alignment of peptide-bound KDELR (rainbow) with Frizzled 4 (PDB: 6BD4, in grey), achieving an RMSD of 10.2 Å over 196 Cα. Rubredoxin fusion is coloured in red and was added to facilitate crystallisation. **B** Structural alignment of peptide-bound KDELR (rainbow) with β2 adrenergic receptor (PDB: 3SN6, in grey), with an RMSD of 13.7 Å over 198 Cα. T4 lysozyme fusion is coloured in red and was added to facilitate crystallisation.

pockets with their C-termini. GLP-1 bound peptide-5 is heavily modified in this structure, but its N-terminal Cap modification is the farthest inside the binding pocket (Jazayeri et al., 2017). Peptide binding of NTS1 is largely hydrophobic in nature, with seven hydrogen bonds, four of which involve back-bone interactions and three hydrogen bonds involving the C-terminal carboxy group with interactions to R327 and Y146, F344 appears to be engaging R8 of the peptide in a cation-π interaction (White et al., 2012). KDELR binding to KDEL peptide involves mostly polar interactions, with 10 KDELR residues providing 14 hydrogen bonds as well as as cation-π interactions and some hydrophobic interactions near the side chain of the terminal leucine.

The binding modes between NTS1 and KDELR are very different in nature, with three exceptions. Firstly, they adopt a similar perpendicular pose in their receptors. Secondly, both engage in cation-π interactions at a relatively solvent-exposed position (F344–R8 in NTS1 and W120–K-4 in KDELR). Cation-π interactions are thought to be very energetically favourable, and relative to salt bridges remain more so in solvent exposed environments due to high the dielectric constant (Gallivan & Dougherty, 2000). Finally, the hydrogen bond coordination of the carboxy of terminal leucine is strikingly similar. In the NT the carboxy group is coordinated by R327 on one side and Y146 on the other, very reminiscent of R159 and Y48 in the KDELR, which has an additional R47 contributing to the interaction.

From the perspective of structural homology it appears that the KDELR, much like SWEETs, is unrelated to GPCRs (Jaehme et al., 2016). Though there are some interesting similarities in how the KDELR and NTS1 recognise their peptides, the general recognition scheme appears to be of a different nature, largely polar in the KDELR and mostly hydrophobic in NTS1. The recognition of small peptides by receptors may indeed follow some common principles due to the nature of the interactions and chemistry possible, an analysis thereof would be of interest from evolutionary and drug design perspective. However, if the KDELR is to activate G-proteins, it is unclear how it interacts with these. An *in vitro* G-protein activation assay with purified components would be an obvious first step in providing direct and convincing evidence of GPCR activity of the KDELR.

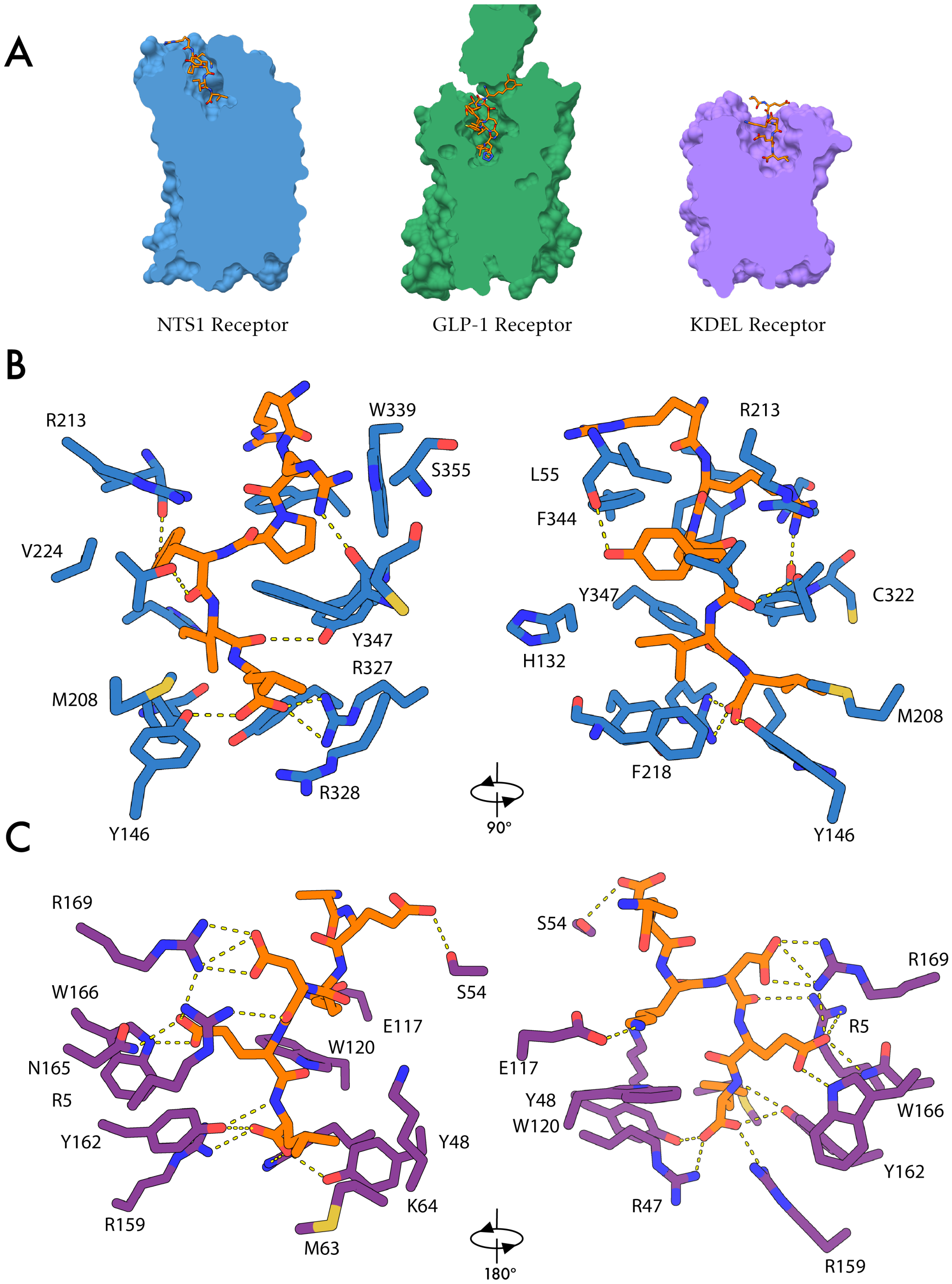


Figure 5.21 **A** Structures of NTS1 receptor (blue, PDB:4GRV (White et al., 2012)), GLP-1 receptor (green, PDB:5NX2 (Jazayeri et al., 2017)) and KDELR (purple, PDB:6I6H) showing the peptide-binding pockets of the receptors. **B** Residues of NTS1 interacting with the NT_{8-13} (orange). **C** Residues of the KDELR interacting with the KDEL peptide (orange).

5.6 Conclusion

The solution of the crystal structure of the KDELR in several different states and the subsequent analysis and comparison of the data lead to several hypotheses regarding the function of the KDELR.

- The KDELR senses pH via a protonation of H12, which, in conjunction with peptide binding, stabilises the peptide-bound state of the KDELR, via a short hydrogen bond.
- The stabilisation of the short hydrogen bond between E127 and Y158 requires the presence of the bound peptide and low pH, however the major conformational changes in the receptor can be induced by the presence of peptide at a neutral pH.
- Peptide recognition occurs in a highly conserved pocket of the KDELR and is sterically constrained to recognition of C-terminal peptides. Any extension of the C-terminal, possibly even by small moieties, may decrease or abolish binding.
- COPI/II binding is regulated by the electrostatic surface properties of the cytosolic side of the KDELR, with COPII machinery being recruited by a negatively charged band and COPI being recruited by the disruption of this band and the formation of a positively charged patch and kink from helix 7. Thus making the recruitment of these COPI and COPII mutually exclusive.
- Syb37 and peptide binding occurs in the same pocket and is thus competitive.
- Structurally the KDELR appears most closely related to transporters, providing support for the initial hypothesis that it may have evolved from an ancestral transporter. Though an evolutionary convergence on the six helical fold cannot be excluded. This question of the KDELR's ancestry and

the role of the PQ-loop would be aided by further study of the PQ-loop family members, both structurally and functionally.

- Structural comparisons to GPCRs appear to show that the KDELR is not a GPCR, though it appears to share some peptide-binding strategies with neurotensin. Hinting at the possibility of some common mechanisms used by receptors to bind and recognise small peptides.

To begin testing the hypotheses generated by the structures of the KDELR, we move to biochemical and cellular assays. Based on the observations in the structures it should be possible to make judicious mutations disrupting the function of the KDELR in a predictable manner, lending evidence to support the hypotheses above. Chapter 6 will explore the KDELR's activity and response to mutations and variation of ligands both *in vivo* and *in vitro*.

6 Molecular Mechanisms

To investigate the hypotheses regarding the molecular mechanisms of the KDELR, generated by the analysis of the crystal structures, a mutagenesis strategy was followed, using two different assays. One *in vitro* radio-labelled peptide binding assay, where the KDELR was in a detergent solubilised form and one *in vivo* retrieval assay, performed in COS7 cells. The assays will be briefly outlined below, followed by the testing of specific hypotheses with the use of these two assays.

6.1 The Assays

Development of an *in vitro* binding assay for the KDELR was a difficult process. The SPR and BLI methods using biotinylated KDELR could not detect the presence of short TAEKDEL peptides and MBP with a C-terminal addition of a KDEL sequence provided no clean signal. Biotinylated peptide could be loaded onto chips but the KDELR bound the chip surfaces non-specifically. MST and FP provided some signal but was very noisy due to the pH sensitivity of the fluorophore attached to the peptide, rendering them inappropriate. GFP-tagged KDELR fluorescence was quenched by KDEL peptides. Finally a ^{3}H-TAEKDEL labelled peptide proved reliable and consistent.

Development of the *in vivo* assay proved somewhat more straight-forward, requiring only optimisation of cell type, promoter and transfection.

6.1.1 *In Vitro* Assay Development

To develop an *in vitro* assay, the initial approach was to develop an assay that would provide rich kinetic information, such as an SPR or BLI binding assay, to interrogate the detailed effects of mutations and pH on the binding process. Some of the troubles encountered in this process are outlined below, in a description of the BLI experiments performed. Given the success with BLI from the Syb37 measurements, biotinylated *Gg* KDELR was immobilised on streptavidin needles as for the sybody measurements. However, the peptides used for crystallisation could not be detected, likely because they were too small. A purified mScarlet-BiP fusion, from the *in vivo* assay below, interferes with the light reflection in BLI and cannot be used. A similar fusion of the C-terminus of BiP to MBP resulted in a protein that could not be detected reliably due to high non-specific binding.

Therefore some biotinylated KDEL peptides were ordered and immobilised on BLI needles. The first row of figure 6.1 shows a peptide concentration scouting experiment, where a range of peptide concentrations has been immobilised on the needles and these are subsequently bound with a constant concentration of KDELR. The curves do not appear to follow a 1:1 binding pattern (the shapes of the curves deviate from 1:1 behaviour) and when we observe some of the raw data on the right, without the reference needle subtracted, we can see strong background binding of the KDELR to the needles. If we perform binding assays the apparent binding affinity is very high. However, the signal at lower concentrations is very small and the reference needles are reporting a similar response level to the peptide-bound needles. This trend worsens with higher pH. The background contribution of non-specific KDELR binding is too high to produce reliable data. Given the low K_d estimates from the 1:1 fitting and the unreliability of the measurements when approaching this value, any fit will be accordingly unreliable. Even binding to DDEL peptides or using mutants with decreased peptide binding affinities (as tested with the radio-labelled peptide binding assay developed later), gave very low K_d estimates, indicating that the binding was not peptide specific.

The non-specific binding of KDELR to SPR and BLI surfaces could not be decreased effectively by changing salt concentrations, detergent concentrations or by the addition of Bovine Serum Albumin (BSA). We are currently looking at returning to the use of biotinylated *Gg* KDELR with native ligands (such as Calreticulin and Cholera A/B toxin), and have procured vectors to produce these, as well as procuring a new range of BLI sensor tips optimised for higher sensitivity and smaller molecule detection, hopefully allowing for more in-depth kinetic analysis of the binding of the KDELR.

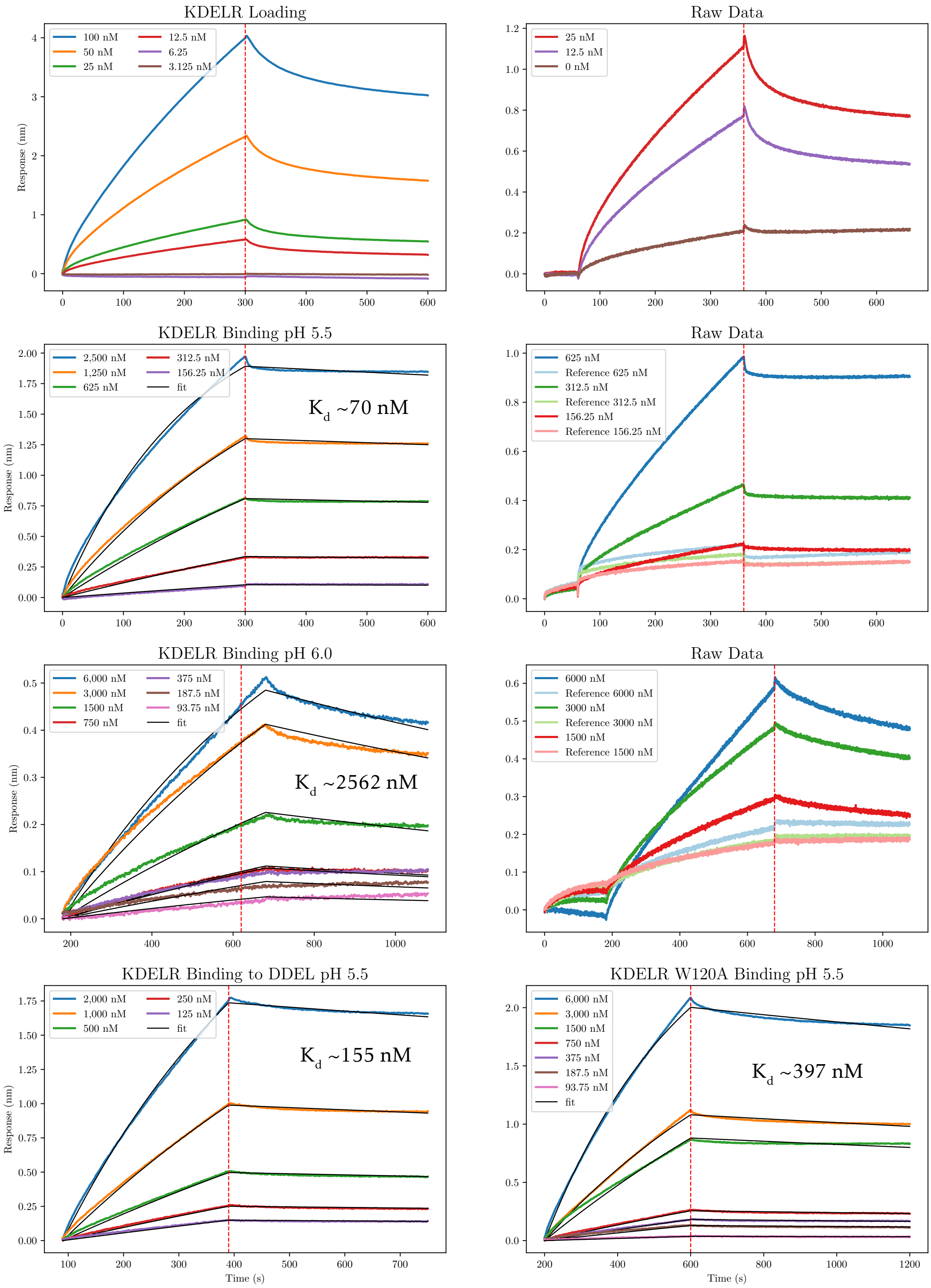


Figure 6.1 BLI experiments performed during assay development. Top left is the response of loading *Gg* KDELR with different amounts of immobilised peptide, reference subtracted. To the right are some of the raw traces as well as the control trace with no immobilised peptide. In the second row we see KDELR binding to immobilised KDEL peptide, with an apparent K_d of 70 nM, to the right some of the raw traces and the reference traces for the corresponding concentration. At lower concentrations the background signal is nearly equal to the signal from the sample. In the next row the same experiment can be seen performed at pH 6.0, where this phenomenon is observed even at higher concentrations of KDELR. Finally, KDELR binding to DDEL and a mutant KDELR (W120A) are shown both with surprising K_d values.

6.1.2 Radio-labelled Peptide Binding

The details of the assay are described in section 8.7.1 and an outline is shown in figure 6.2. Generally assays were performed in a low salt buffer, with detergents present. KDELR was incubated with radio-labelled peptide (specific activity of 106 Ci $mmol^{-1}$ for ^{3}H-TAEHDEL and 128 Ci $mmol^{-1}$ for ^{3}H-TAEKDEL) at room temperature and then bound to a mixed cellulose ester filter and washed. Bound peptide was quantified by scintillation counting. Each concentration series or measurement was repeated at least three times. TAEKDEL represents the last seven amino acids of human BiP (*Gg* BiP has the sequence AAEKDEL), a known KDELR ligand. NanoDSF results were used to refine the buffer conditions of the experiment, which indicated that increased salt concentrations decreased the stabilisation of peptide binding (See Appendix B, figure B.5). Mixed cellulose ester filters provided the highest binding and two 500 µL washes of buffer provided the minimal washing required to remove background binding to the filter. This assay thus allows a simple readout of the affinity of the KDELR and KDELR mutants to the peptide as well as the IC_{50} of other non-tritiated peptides in competition with ^{3}H labelled peptides.

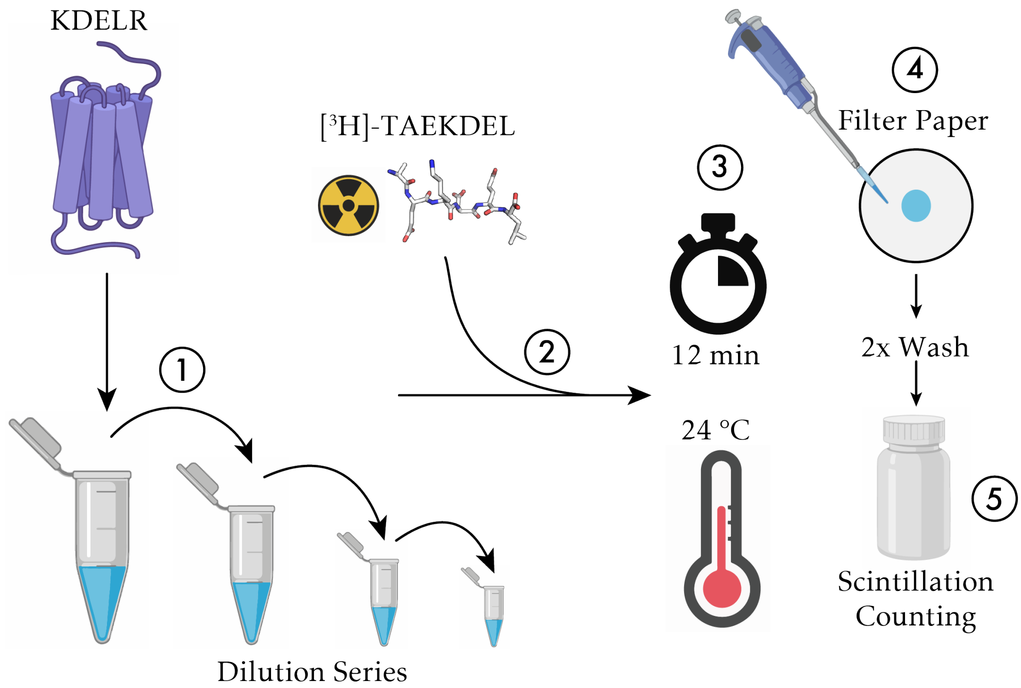


Figure 6.2 1 A dilution series of KDELR is prepared and 2–3 incubated with radiolabelled peptide. 4 The mixture is placed on filter paper on a vacuum manifold, where it is washed twice with buffer. 5 The washed filter paper is placed in a scintillation vial for scintillation counting.

6.1.3 Cellular Retrieval Assay

The details of this assay are described in section 8.8 and an overview is seen in figure 6.3. COS7 cells were transfected with KDELR-GFP fusion and mScarlet-KDEL (artificial ligand, with the 16 C-terminal residues of human BiP) or empty vector (no ligand). The KDELR localises to the Golgi under steady-state conditions, however under a large excess of ligand, the KDELR will relocalise to the ER, thus providing a readout for ligand sensitivity of the KDELR *in vivo* (Lewis & Pelham, 1992a). COS7 cells were chosen for their tight Golgi distribution, allowing for easier quantification. Human elongation factor-1 α (EF1α) promoters were used as they provided lower expression than cytomegalovirus (CMV) promoters, where many cells were highly over-expressing proteins, leading to signal everywhere in the cells, making them unusable. Transfection amounts were optimised to maximise the number of cells expressing the minimal amounts of both proteins for efficient detection, avoiding gross over-expression and any

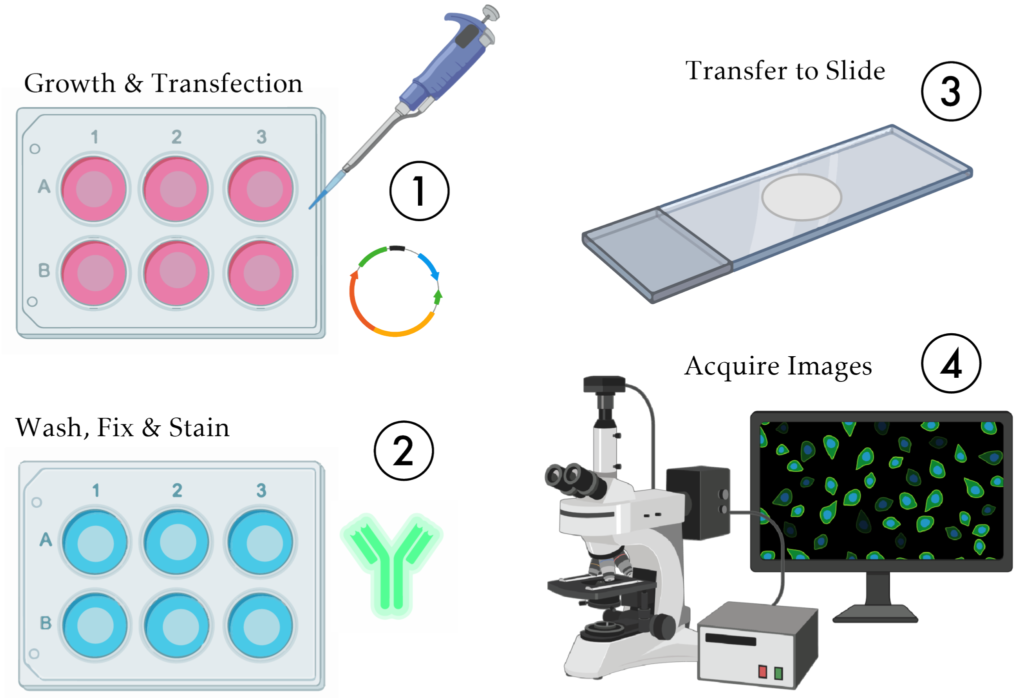


Figure 6.3 1 COS7 cells growing on glass coverslips in 6 well trays are transfected with KDELR and ligand. 2 Media is removed, cells are washed, fixed, permeabilised and stained. 3 The coverslips were transferred to glass slides and 4 images were taken in a fluorescence microscope.

unwanted effects caused by this. This assay provides an *in vivo* complement to the *in vitro* assay above, allowing for greater confidence in the results obtained.

6.1.4 *In Vitro* Mutant Validation

In order to ensure that KDELR mutants that were purified showed any differences in binding due to their specific effects, rather than a lack of structure of the KDELR, CD spectra were measured to check for secondary structure and CD melts were subsequently performed to check for structural stability. The results of these experiments are shown in figure 6.4, SDS-PAGE runs and CD melts are seen in figure 6.5 and 6.6. R159 mutants could not be purified to sufficient quantities to perform these or any other experiments, as they precipitated heavily during the TEV cleavage stage of purification. All remaining mutants show secondary structure, with varying levels of stability, indicating that some mutations may have a slight destabilising effect on the structural integrity of the KDELR.

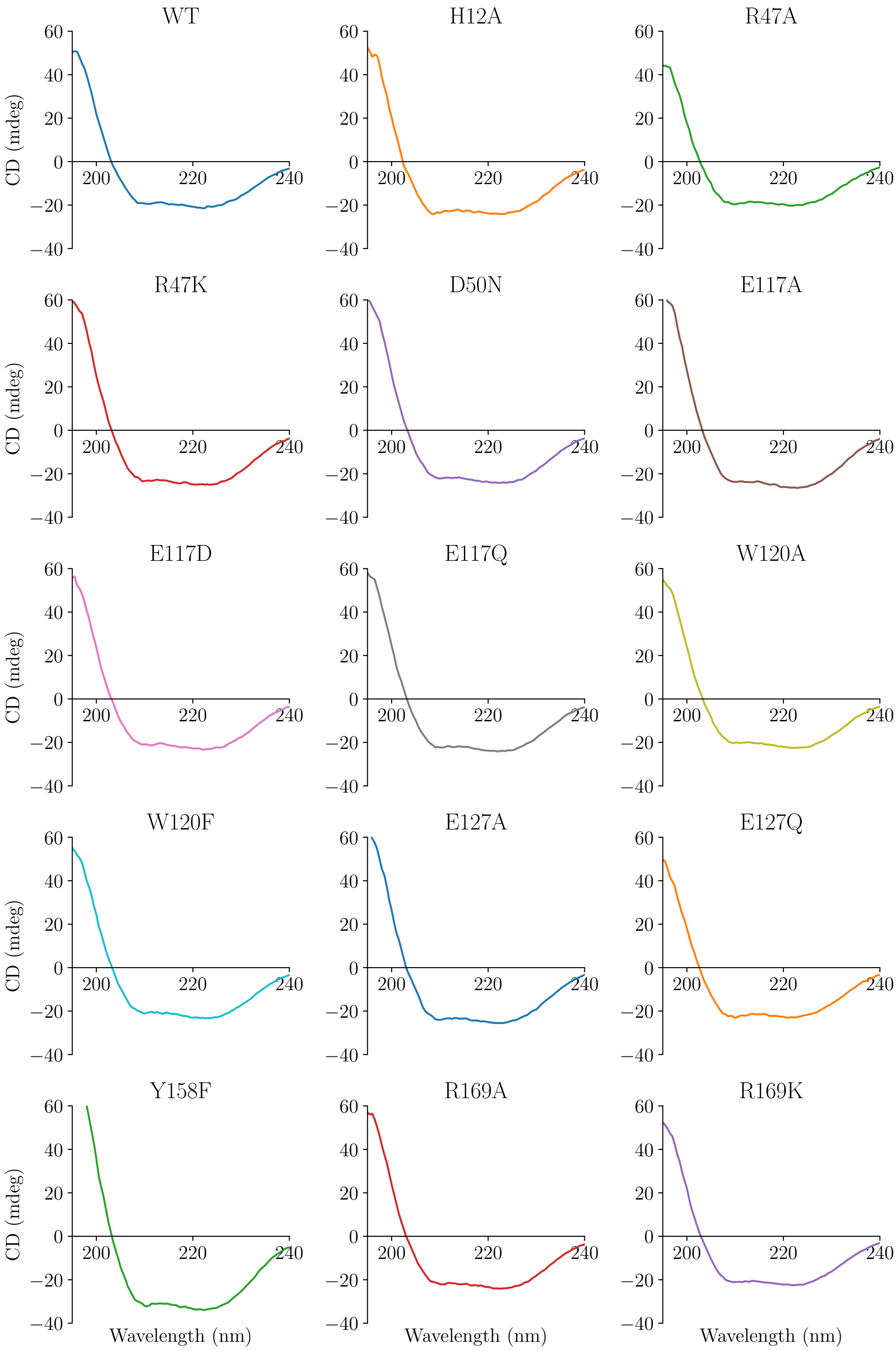


Figure 6.4 CD spectra of *Gg* KDELR and mutants, in 10 mM potassium phosphate pH 6.5, 50 mM Na_2SO_4 and 0.03% DDM 0.0015% CHS.

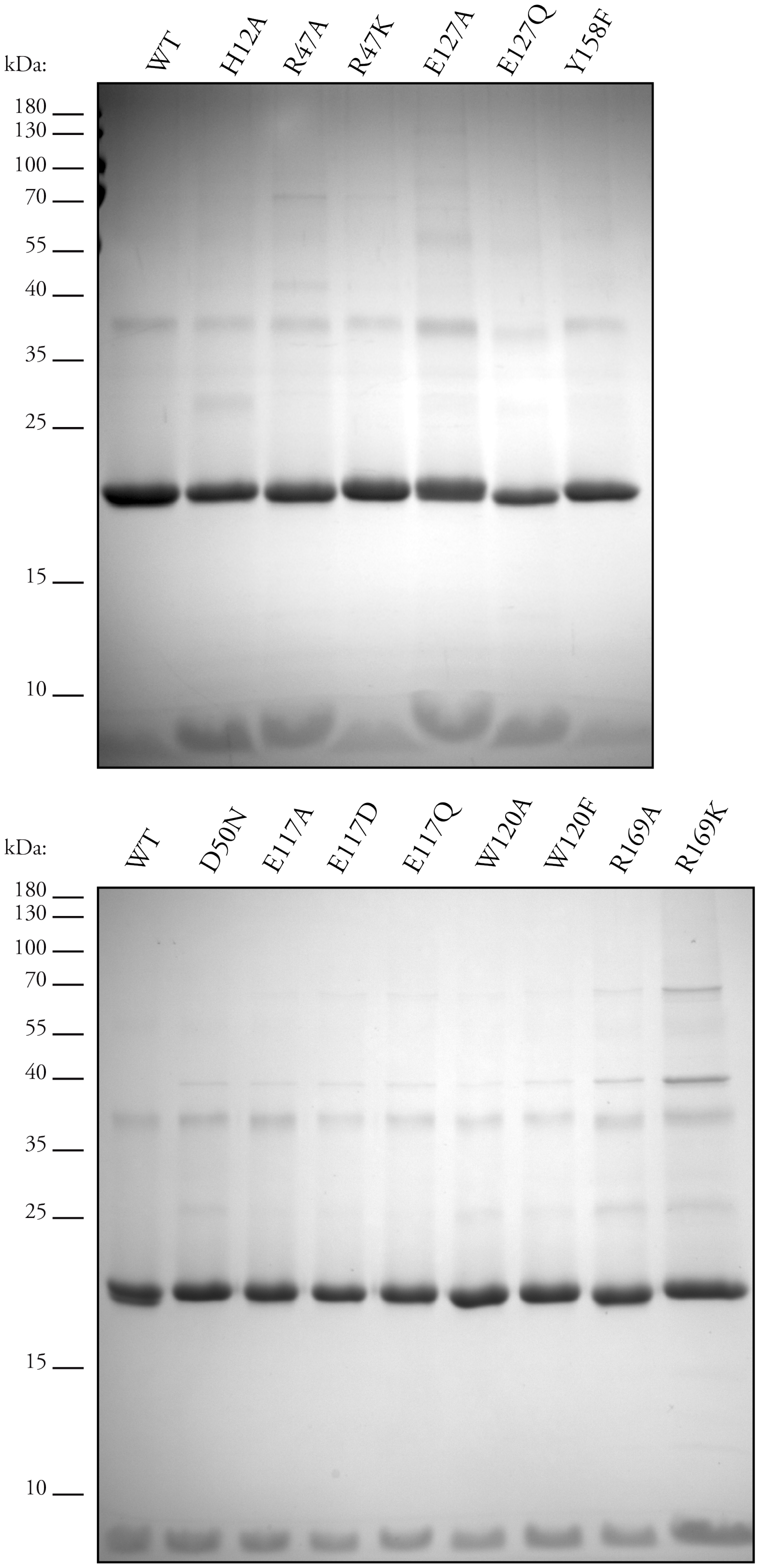


Figure 6.5 SDS-PAGE of all *Gg* KDELR mutants used for radio-labelled *in vitro* assay. Stained with coomassie brilliant blue.

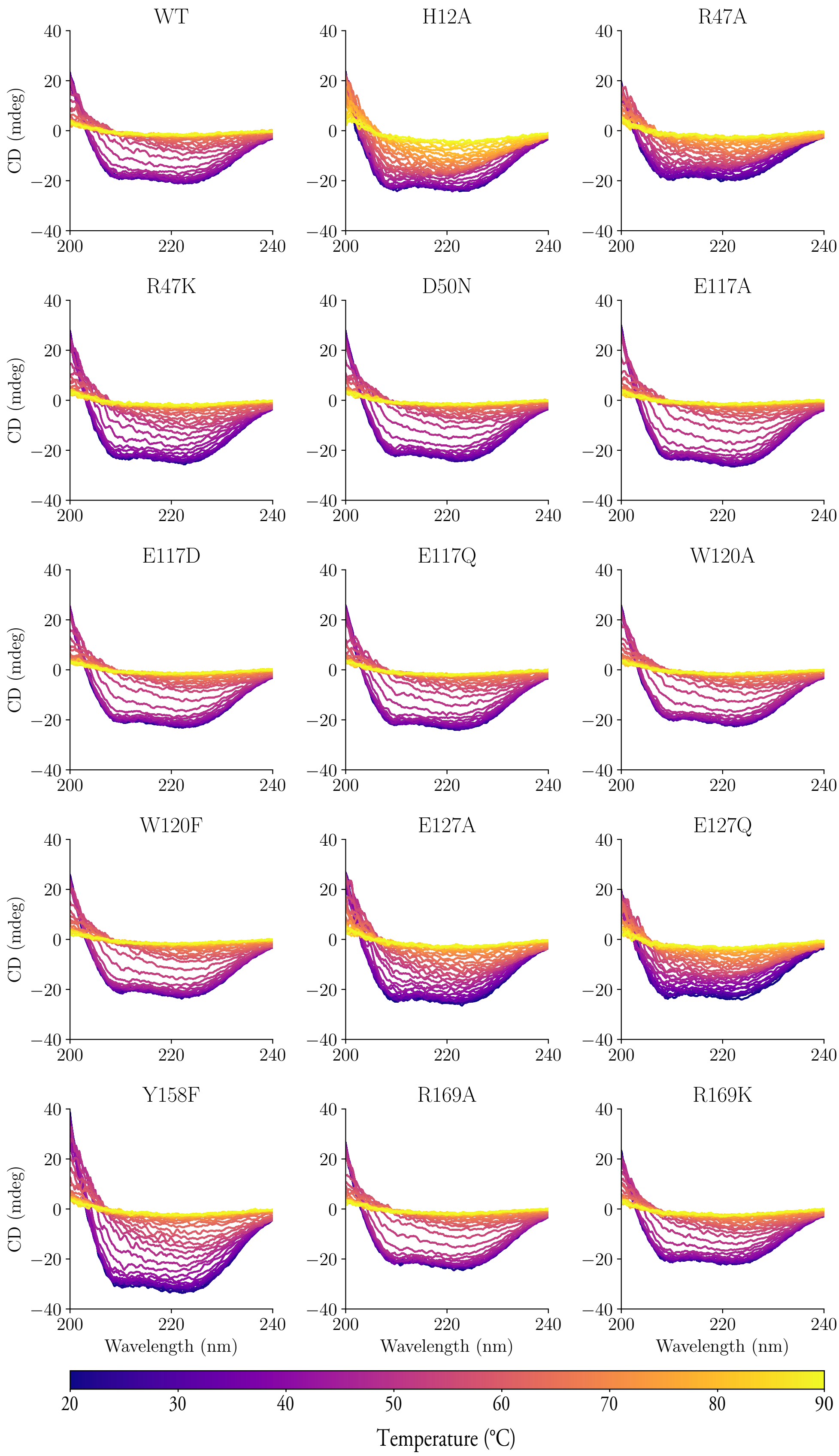


Figure 6.6 CD melts of the *Gg* KDELR and mutants, in 10 mM potassium phosphate pH 6.5, 50 mM Na_2SO_4 and 0.03% DDM 0.0015% CHS.

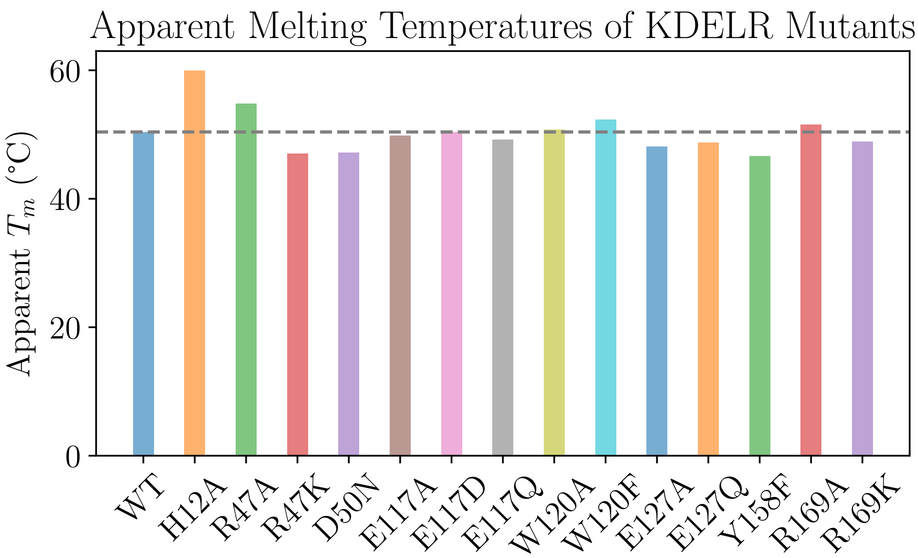


Figure 6.7 Apparent melting temperatures of the *Gg* KDELR and mutants, at 210 nm. The dashed line indicates the WT T_m. The full melting profiles can be seen in supplementary figure B.7.

6.2 pH-Dependent Binding *In Vitro*

The pH-dependence of peptide binding was first discovered by **Wilson et al.** (1993), with an optimum of around pH 5. Testing the binding of KDEL and HDEL peptides, a pH-dependence was also observed in the detergent solubilised KDELR, see figure 6.8 and table 6.1. At a pH lower than 5, KDELR began visibly precipitating during the course of the assay so no lower pH values were measured. Demonstration of the pH-dependent binding of the KDEL and HDEL peptides provides a validation for the assay. Of particular interest in these assays is the observed difference in affinity of the *Gg* KDELR2 to TAEKDEL and TAEHDEL peptides, with a tenfold higher affinity for HDEL at pH 5.4. This comes as a surprise as the ER proteins in *Gg* generally are found to have KDEL and not HDEL C-terminal motifs. Our initial search had shown no Uniprot hits for HDEL C-terminal proteins in *Gg*, against five for KDEL. Searching Uniprot but now including TrEMBL for the ER retention motif with the PROSITE pattern PS00014 ([KRHQSA]-[DENQ]-E-L>) in *Gg* we get 76 hits, searching only for K-[DENQ]-E-L> we get 29 hits and 20 for H-[DENQ]-E-L> (**UniProt Consortium**, 2019; **de Castro et al.**, 2006). Similar results are obtained if we search *Gg* in the Kyoto Encyclopedia of Genes and Genomes database (**Kanehisa & Goto**, 2000). Chicken is known to code for two KDELRs, KDELR2 and KDELR3 (identified

	Apparent K_d (μM ± SD)	
pH	KDEL	HDEL
5.0	1.25 ± 0.14	NA
5.4	1.87 ± 0.46	0.26 ± 0.04
5.9	3.93 ± 0.50	1.13 ± 0.12
6.4	7.46 ± 2.89	1.98 ± 0.14

Table 6.1 Apparent K_d of peptide (TAE[HK]DEL) binding to *Gg* KDELR at various pH, Experiments were performed three times, each measurement performed once. Peptide was present at 10 nM with 20 mM buffer (MES or Tris-HCl), 40 mM NaCl and 0.01 % DDM:CHS (20:1).

from computational annotation) which share 75% identity. It is conceivable that these two receptors have differential KDEL motif preferences, much like has been observed for the three human KDELRs Raykhel et al. (2007).

Previously the affinity of *Hs* KDELR1-myc-his to ^{125}I-labeled YTSEKDEL peptide, was determined to be between 45–167 nM, depending on the lipid environment (**Scheel & Pelham**, 1996). These results are in a similar range to those measured for *Gg* KDELR2 binding to HDEL. There have been indications that *Hs* KDELR paralogues may have preferences for the -4 position and possibly the -5 and -6 positions too (**Wilson et al.**, 1993; **Raykhel et al.**, 2007; **Mei et al.**, 2017; **Alanen et al.**, 2011). Investigation of molecular recognition of the four C-terminal residues will be discussed in section 6.5 and a comparison of H/KDEL binding to the KDELR has been made in section 5.2.4.

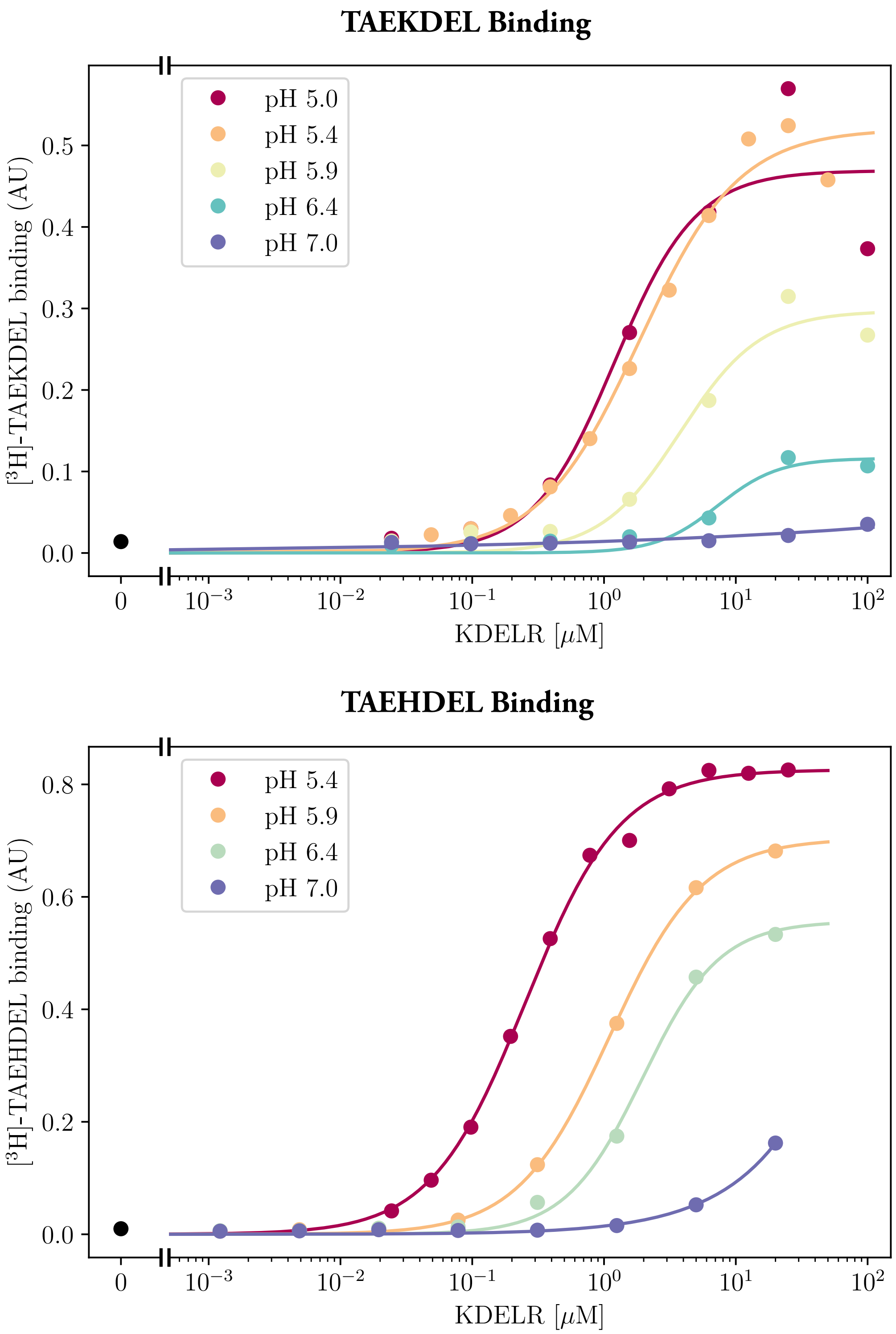


Figure 6.8 Results of *in vitro* binding assay in response to variation of pH, showing representative curves of single experiment, which were repeated three times. Peptide present at 10 nM, in 20 mM MES pH 5.4, 40 mM NaCl, 0.01% DDM:CHS (20:1).

6.3 Cellular Retrieval

The cellular retrieval assay was benchmarked with the wild type (WT) *Hs* KDELR1, with and without an artificial ligand. As seen in figure 6.9, without ligand, KDELR localises to the Golgi, upon addition of the KDEL ligand, the KDELR relocalises to the ER. To quantify this effect, the fraction of fluorescent signal from the KDELR in the Golgi was calculated by defining the location of the Golgi using TGN46 and then calculating the background subtracted fraction of KDELR signal present in the Golgi as fraction of the total signal of the cell, see section 8.8.4 for details. This experiment validated the assay parameters, and provided a benchmark for the retrieval effect with WT protein.

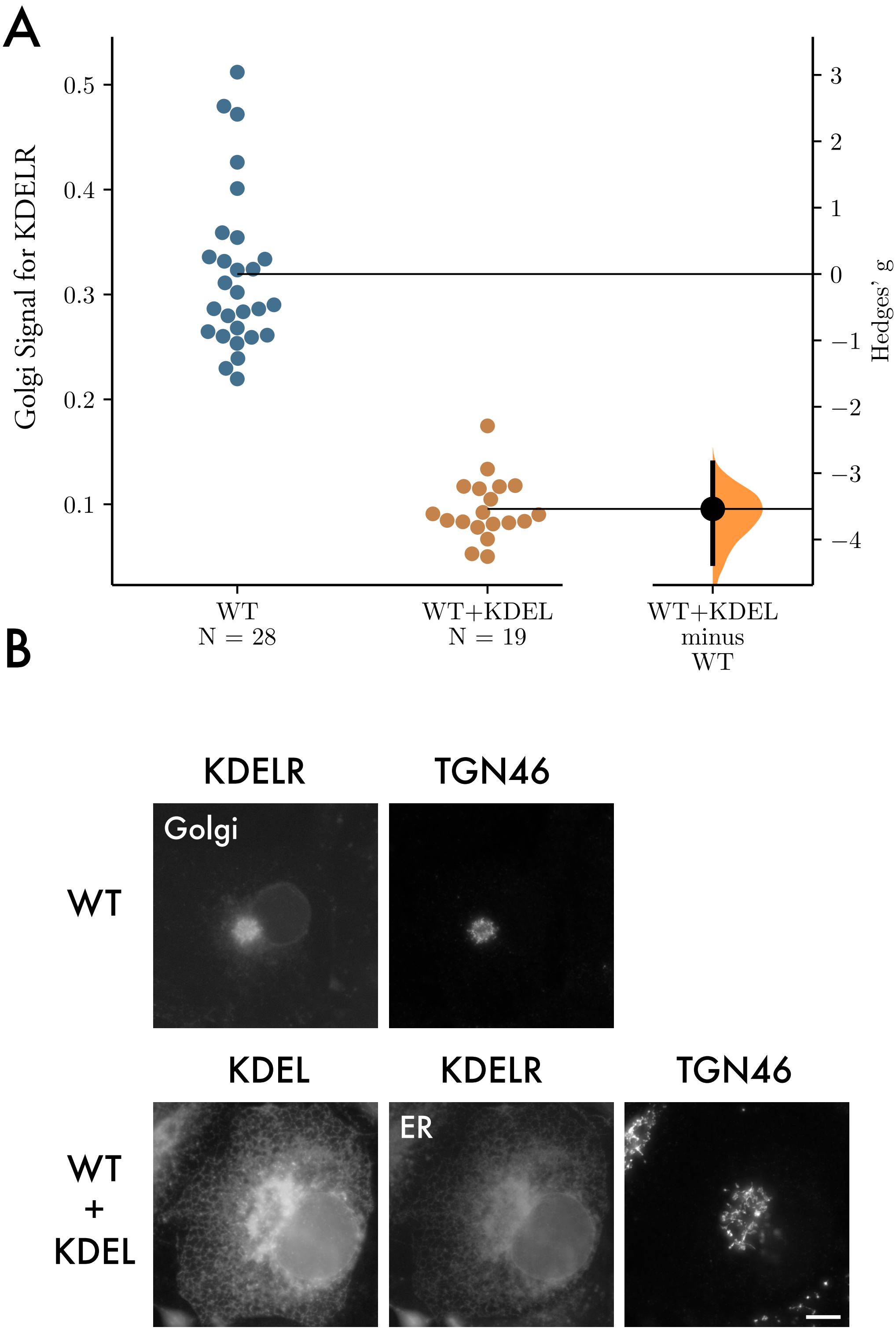


Figure 6.9 ER retrieval assays were performed in the absence or presence (+) of KDEL ligand and the Golgi signal for KDELR is plotted in **A** as a Gardner-Altman plot, displaying the effect size as a bootstrap 95% confidence interval below (Ho et al., 2019). **B** KDELR was tested for KDEL ligand-induced redistribution from Golgi to ER as in **A**. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

6.4 pH-dependence Mechanism

To collect evidence for the short hydrogen bond hypothesis, wherein protonation of H12 and presence of peptide stabilise the bound state, several mutations were introduced to disrupt this process. H12A, E127A, E127Q and Y158F directly disrupt the three residues critical in this process. R47A, R47K, R159A and R159K disrupt the connection between the C-terminal carboxy of the peptide, decoupling the proton sensing H12 from the peptide binding pocket (refer to figure 5.16).

6.4.1 *In Vitro* Binding Assays

Figure 6.10 shows the results of the binding assays with these mutants with both TAEKDEL and TAEHDEL peptides. In the case of TAEKDEL, all mutations greatly reduce binding. For TAEHDEL, which has a higher affinity in the WT receptor, the picture is somewhat more subtle. E127 mutations appear to bind at 25–40% of WT, with a polar Q being favoured over the more extreme A mutation. Y158F surprisingly is able to bind at half of WT. Both R47 mutations reduce binding to about 40%. R159 mutants could not be purified, indicating R159's role in the structural integrity of the KDELR. Interestingly, H12A makes no direct contacts to the bound KDEL peptide, yet completely abolishes binding in the assay.

C-terminal Amidation

To further investigate the role of the C-terminal carboxy group of the peptide in this process, TAEKDEL peptides were ordered with a C-terminal amidation, masking the charged carboxy group. IC_{50} curves comparing the amidated peptides to unmodified TAEKDEL and TAEHDEL peptides, can be seen in figure 6.11. The C-terminal amidation causes a large disruption to binding, either through steric interference, the loss of the charge interactions or, most likely, both. From Prosite motif searches, no KDEL motifs with a further C-terminal residue after the leucine were found in the swissprot database of chicken. Due to the amino acid modification, this experiment could not be easily replicated in the cellular context.

6.4.2 Cellular Retrieval Assay

Figure 6.12 shows the the results of the cellular retrieval assay on the KDELR mutants designed to disrupt the short hydrogen bond. Images of each mutant in cells are shown in figure 6.13. All mutants show a loss of sensitivity to the presence of ligand and do not relocate. R159 mutants fail to reach the Golgi and localise to the ER both with and without ligand stimulation, indicating some trafficking defect, possibly due to structural instability, consistent with the precipitation observed when purifying these mutants. The *in vivo* data complements the *in vitro* experiment and they appear to validate the results.

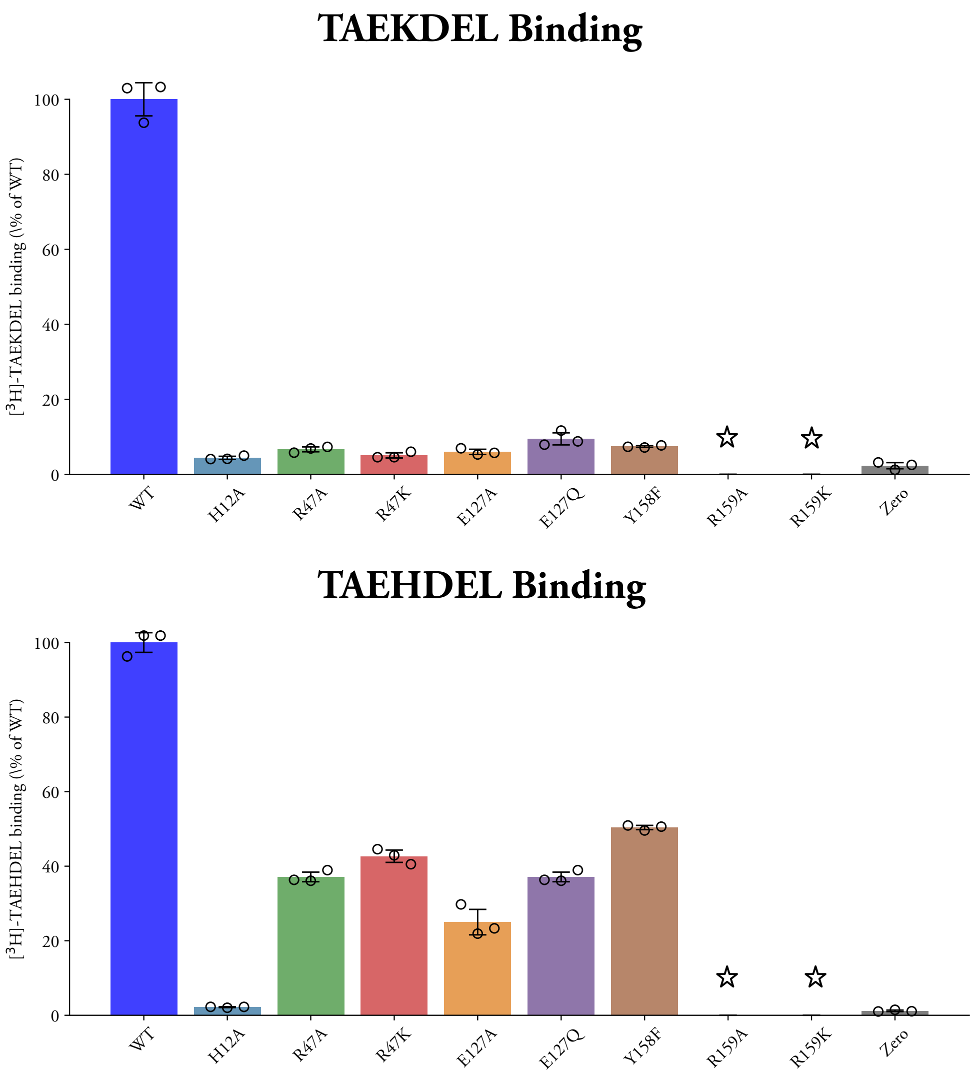


Figure 6.10 Results of the *in vitro* binding assay, with ^{3}H-TAE[KH]DEL peptides to the KDELR and KDELR mutants. Stars indicate mutants that could not be purified. Peptides were present at 10 nM, protein was at 25 µM, in 20 mM MES pH 5.4, 40 mM NaCl, 0.01% DDM:CHS (20:1). Experiments were performed three times, measured once, error bars are SD.

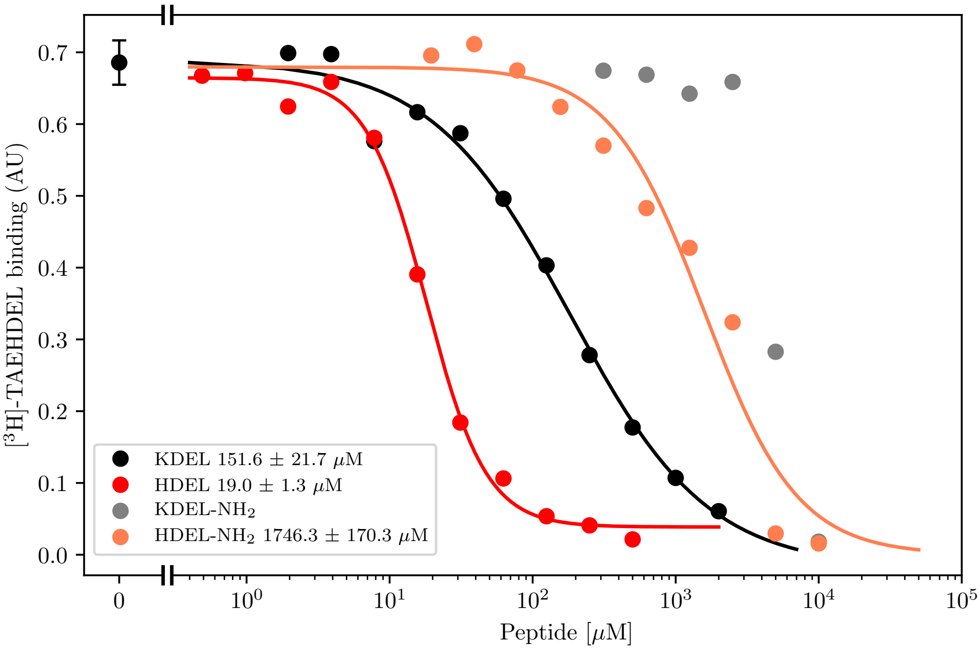


Figure 6.11 Results of competition assays against ^{3}H-TAEHDEL, showing representative curves, experiments were repeated three times and measured once. ^{3}H-TAEHDEL present at 10 nM, protein was at 7 μM, in 20 mM MES pH 5.4, 40 mM NaCl, 0.01% DDM:CHS (20:1).

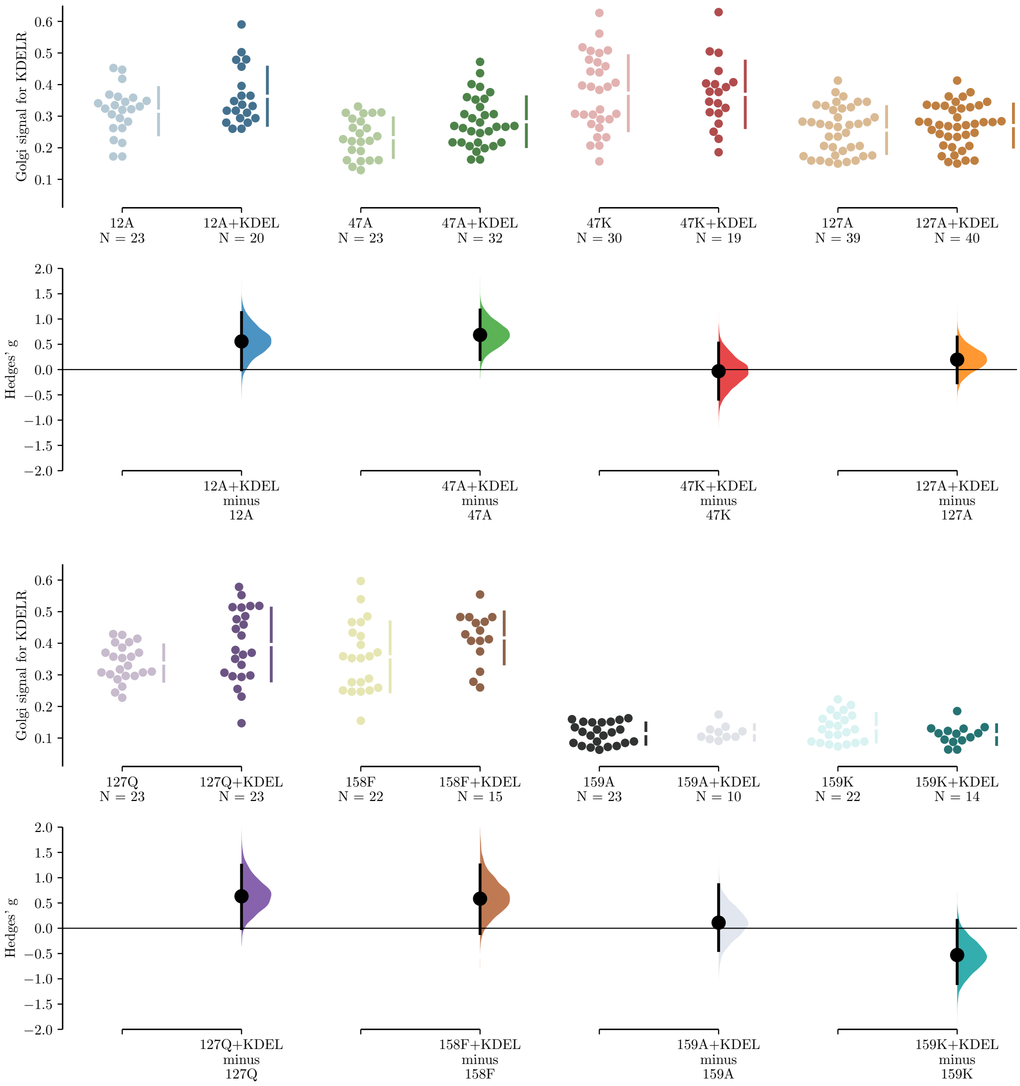


Figure 6.12 ER retrieval assays were performed in the absence or presence (+) of KDEL ligand and the Golgi signal for KDELR is plotted as a Cumming plot, displaying the effect size as a bootstrap 95% confidence interval below and the standard deviation as a vertical error bar to the right of the swarm plot, with the mean represented as a gap (Ho et al., 2019).

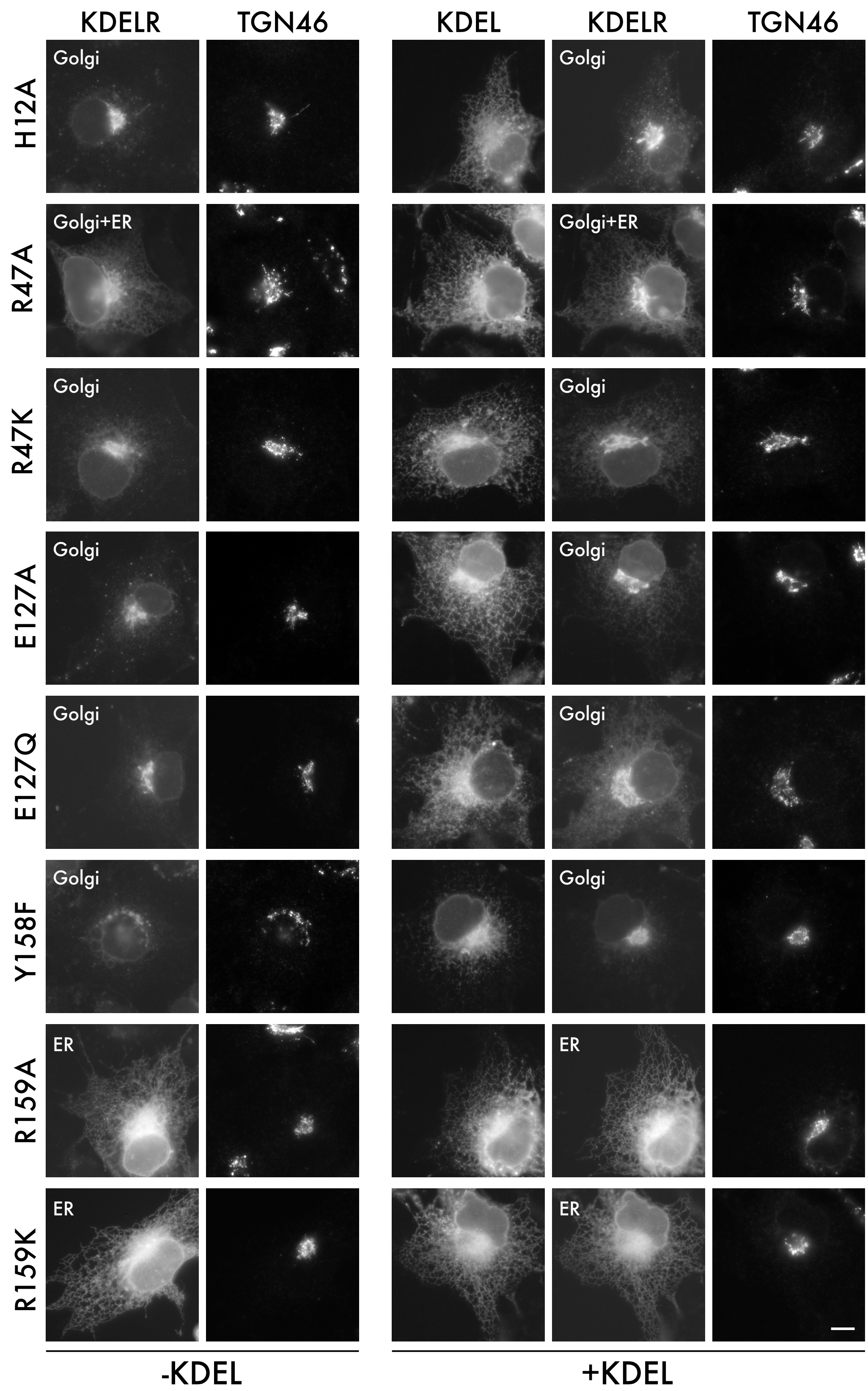


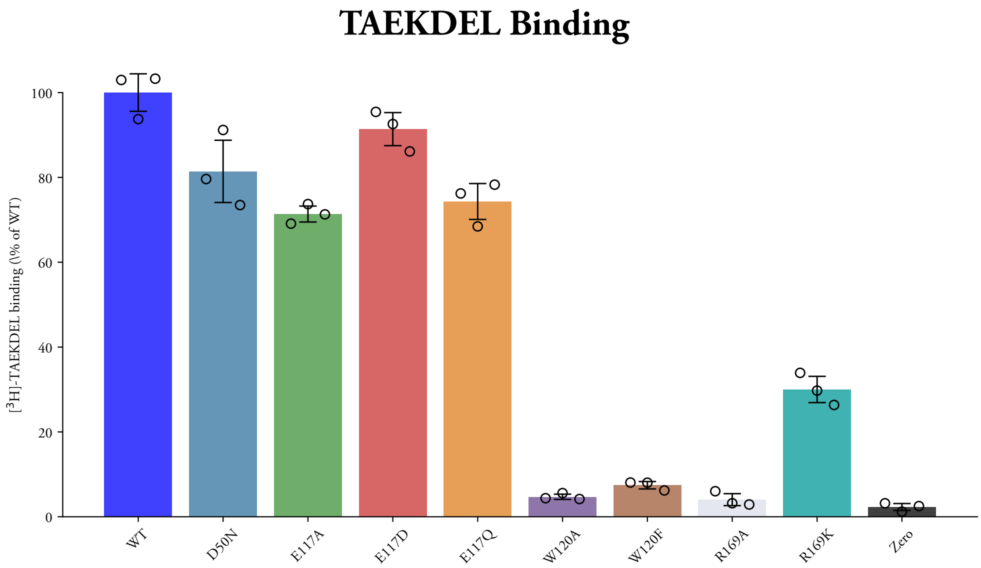
Figure 6.13 Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and an artificial KDEL ligand (termed KDEL). KDELR mutants were tested for KDEL ligand-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

6.5 Peptide Binding & Specificity

To begin investigations of the peptide binding properties and preferences of the KDELR, a combination of *in vitro* and *in vivo* strategies was taken. *In vitro*, mutations were made to disrupt the binding pocket of the KDELR and different KDEL peptides were used to compete against ^{3}H-TAEHDEL peptides. *In vivo*, these experiments were replicated by using different C-terminal KDEL motifs with WT and mutant KDELR. This data should indicate if the selected mutants in the binding pocket significantly affect peptide binding and the specificity for a motif. As the structural data shows both H-4 and K of the KDEL peptide interact with W120 and E117. These two residues were mutated to W120A, W120F and E117D, E117Q and E117A. R169 forms a salt bridge with the D-3 of the peptide, so this too was mutated to R169A and R169K. D50N was also mutated as it had been observed in the literature that this residue may be important for DDEL specificity in *Kluyveromyces lactis* (*Kl*), and D50C appears to decrease HDEL, KDEL and RDEL binding with no effect on DDEL binding (**Lewis & Pelham, 1992a**; **Semenza & Pelham, 1992**; **Scheel & Pelham, 1998**).

6.5.1 *In Vitro* Competition & Binding Assays

Binding to a HDEL and KDEL peptide, D50N showed near WT binding, indicating a weak effect. Previous observations of no effect on DDEL binding may be due to the loss of the negative charge on D50, which would be ~ 4.4 Å away from the negatively charged D-4 on the peptide, when mutated on the TAEKDEL structure. This may compensate for the slight apparent loss in binding to KDEL and HDEL variants. In homology models of *Kl* ERD2 (52% identity with *Gg* KDELR, figure 6.15), the two residues that appear different in the peptide-binding pocket near the -4 position of the peptide are D50 (in *Gg* KDELR numbering or N51 in *Kl*) and E117, which becomes Q117 in *Kl* (Q123), removing another negative charge around the -4 position. Additionally, F175 is an R (R181), adding a positive charge near the D-3, which may even form a salt bridge. A triple



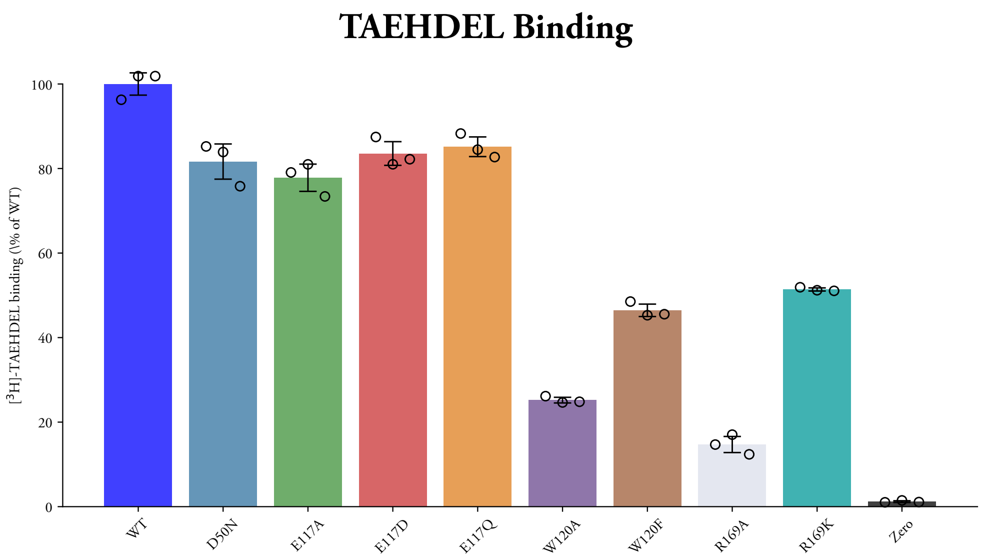


Figure 6.14 Results of *in vitro* binding assay with ^{3}H-TAE[KH]DEL peptides to the KDELR and mutants thereof. ^{3}H-TAEHDEL present at 10 nM, protein was at 25 µM, in 20 mM MES pH 5.4, 40 mM NaCl, 0.01% DDM:CHS (20:1). Experiments were performed three times, measured once, error bars are SD.

mutant (D50N, E117Q, F175R), may be able to increase *Gg* KDELR affinity for DDEL. In *Kl* we also observe S54K and D112R on the solvent-exposed edge of the peptide-binding site. These may potentially engage with segments of the peptide beyond the final four resides.

E117 mutants all showed similar amounts of binding to D50N when binding to HDEL peptide but E117D, which retains the negative charge, improves binding to a KDEL peptide as compared to E117A or Q. W120A and F show no binding to KDEL peptides but some some residual binding is seen in W120A to HDEL, which is improved when mutated to W120F, indicating the presence of an aromatic interaction between the H-4 of the peptide and W120. R169A shows no binding to KDEL and very weak binding to HDEL. This is improved when the charge is reintroduced, and R169K rescues some binding in both cases.

Competition of TAE[HKDR]EL peptides against ^{3}H-TAEHDEL can be seen in figure 6.16. HDEL shows the highest affinity, with R and K being effectively the same, and DDEL binding poorly. Indicating that a positive -4 position increases binding affinity and that the aromatic contribution from H-4 to W120 is significant.

The combination of cation-π and π-π interaction between H-4 and W120 may account for the observed difference in affinity of HDEL vs KDEL peptides, likely by slowing the off rate. Charged His^{+}-Trp interactions have been calculated to be very favourable (**Liao et al., 2013**). In fact, 26% of all tryptophan residues in the PDB are involved in energetically significant cation-π interactions (**Gallivan & Dougherty, 1999**). Collaborators are currently performing calculations to explore the comparative energetic contributions of histidine (protonated and de-protonated) and lysine with W120.

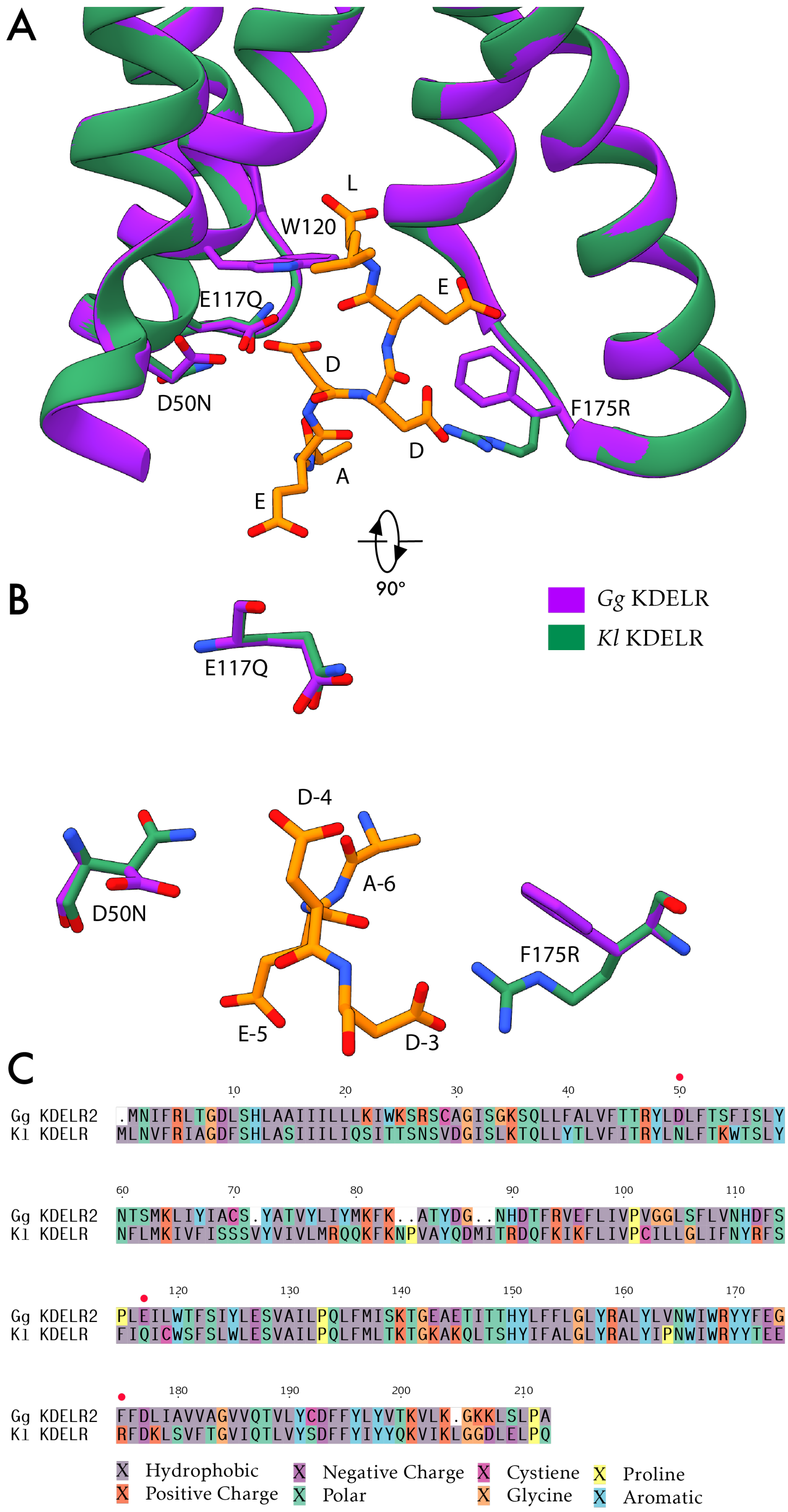


Figure 6.15 **A** Homology of the *Kl* KDELR, based on the KDEL bound *Gg* KDELR structure. The homology model was created with SWISS-MODEL, and manually curated in UCSF Chimera (H-4D was mutated in chimera and the F175R rotamer was selected based on the rotamer library from Shapovalov & Dunbrack (2011)) (Waterhouse et al., 2018; Pettersen et al., 2004). **B** View of the DDEL peptide from top down, showing only the -6–3 residues (AEDD) of the peptide and removing the cartoon representation of the KDELR. **C** Sequence alignment of *Gg* KDELR2 with *Kl* KDELR. Red dots indicate the predicted critical mutations for DDEL recognition.

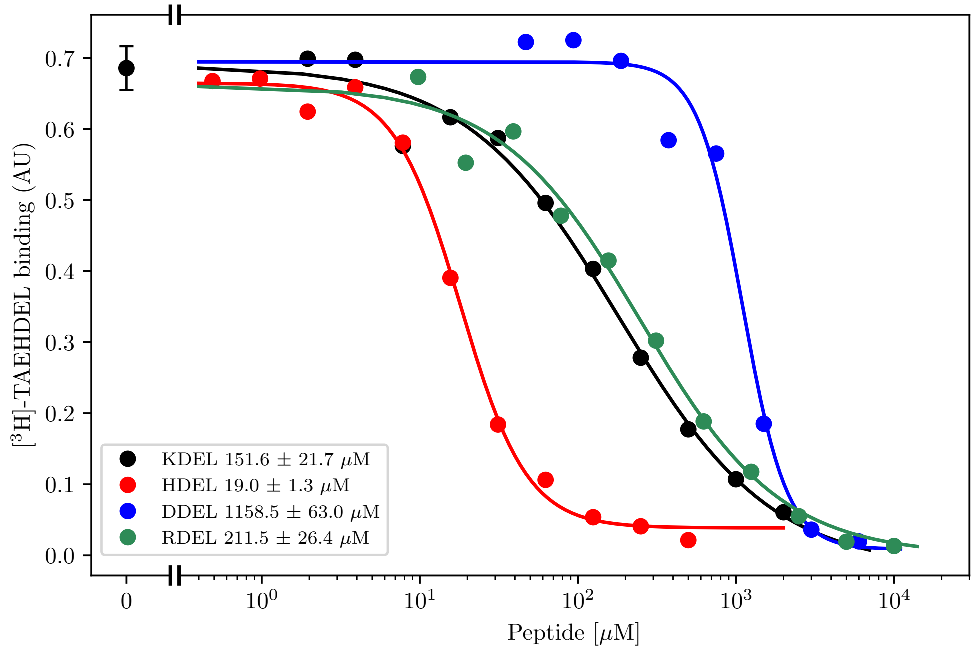


Figure 6.16 Results of competition assays against ^{3}H-TAEHDEL, showing representative curves, experiments were repeated three times and measured once. ^{3}H-TAEHDEL present at 10 nM, protein was at 7 µM, in 20 mM MES pH 5.4, 40 mM NaCl, 0.01% DDM:CHS (20:1).

6.5.2 Cellular Retrieval Assay

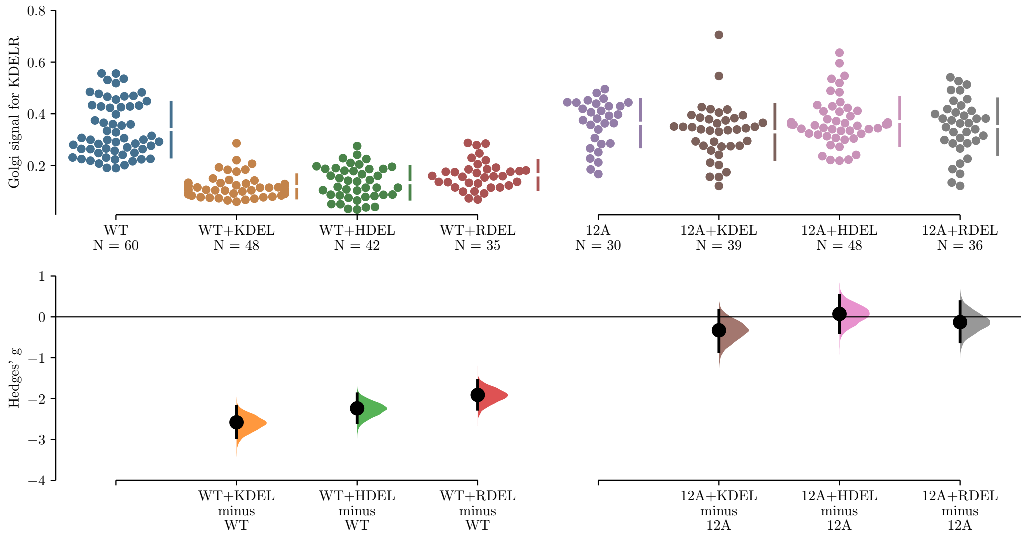


Figure 6.17 ER retrieval assays were performed in the absence or presence (+) of K/H/RDEL ligand and the Golgi signal for KDELR is plotted as a Cumming plot, displaying the effect size as a bootstrap 95% confidence interval below and the standard deviation as a vertical error bar to the right of the swarm plot, with the mean represented as a gap (Ho et al., 2019).

Figure 6.17 shows the results of the *in vivo* assay for WT KDELR probed against KDEL, HDEL and RDEL ligands, all three of which are able to induce a relocation from Golgi to ER. As a control experiment H12A, which does not retrieve ligand, was tested against HDEL and RDEL ligands, which it also fails to retrieve. Figure 6.18 shows that the E117 mutants are generally retrieved, though E117Q appears to be favoured slightly by an RDEL ligand. HDEL ligands appear to be the only ones to be able to induce retrieval of W120 mutants, with a slight effect on both. R169A also appears to have a small retrieval effect with HDEL, while R169K shows some retrieval with all ligands. Examples of cells for each condition can be seen in figure 6.19.

Supplementary figure B.8 shows a selection of cells where the Golgi signal is plotted against the ligand intensity, to observe if relocalisation is induced by higher ligand expression levels. W120 mutants show some sensitivity to HDEL ligand concentration and R169K mutants appear slightly sensitive to ligand expression levels. WT and H12A appear totally insensitive to ligand expression levels. It must be noted that the cells are not depleted of endogenous KDELR, so any effects this may have are not accounted for (e.g. retrieval of ligand and potential oligomerisation with mutant KDELR.)

It appears the cells are less sensitive to KDEL motif variation than the *in vitro* assay, where a nearly ten fold difference in IC_{50} is observed between HDEL and KDEL or RDEL motifs. All three sequences are retrieved by WT KDEL. The charge on R169 appears equally important for all three ligands, though HDEL produces a small effect in the presence of R169A. The interaction of H-4 with W120 appears to be confirmed by the cellular data as the HDEL ligand was the only one to induce some redistribution with the W120F mutant.

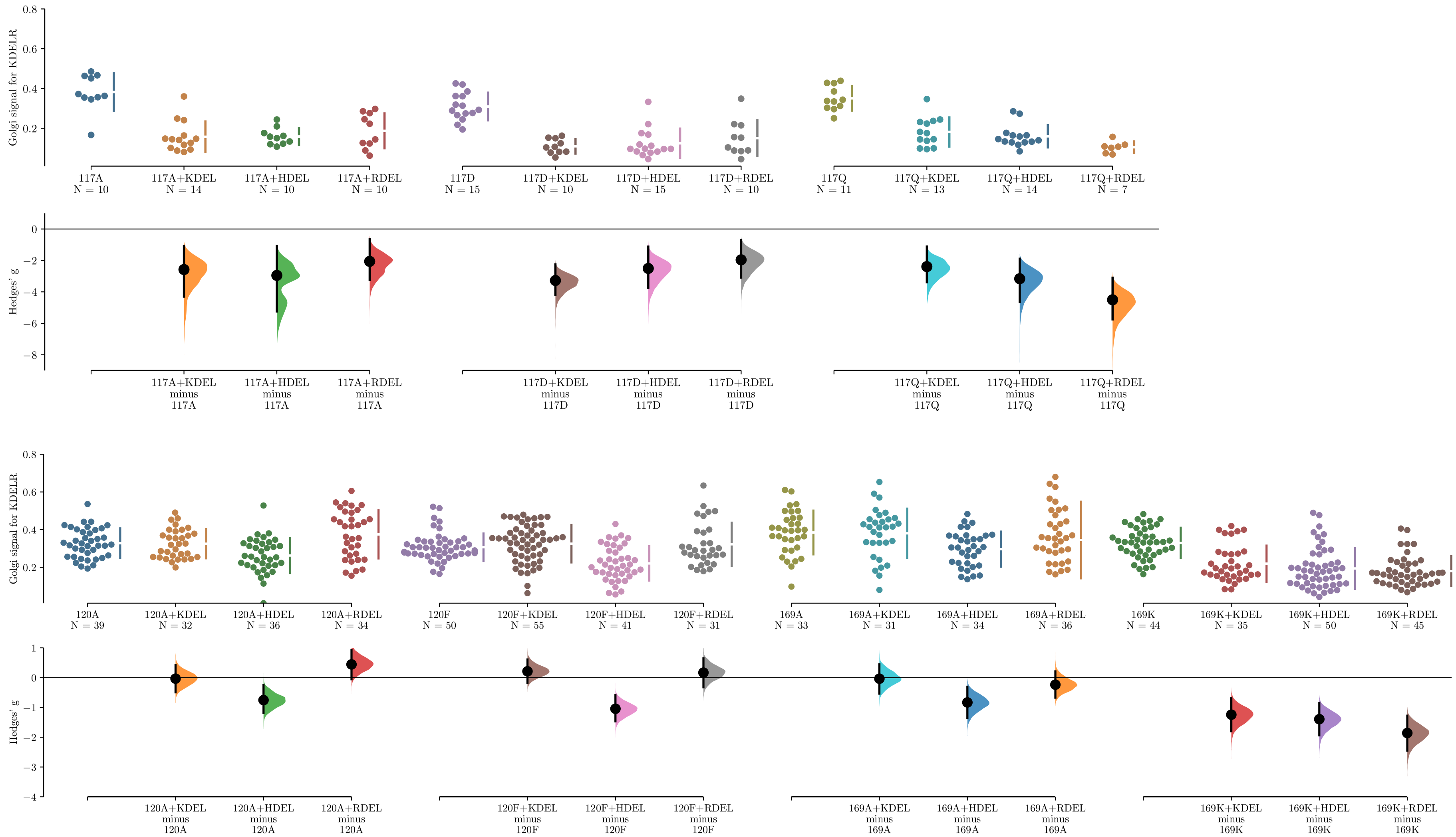


Figure 6.18 ER retrieval assays were performed in the absence or presence (+) of K/H/RDEL ligand and the Golgi signal for KDELR is plotted as a Cumming plot, displaying the effect size as a bootstrap 95% confidence interval below and the standard deviation as a vertical error bar to the right of the swarm plot, with the mean represented as a gap (Ho et al., 2019).

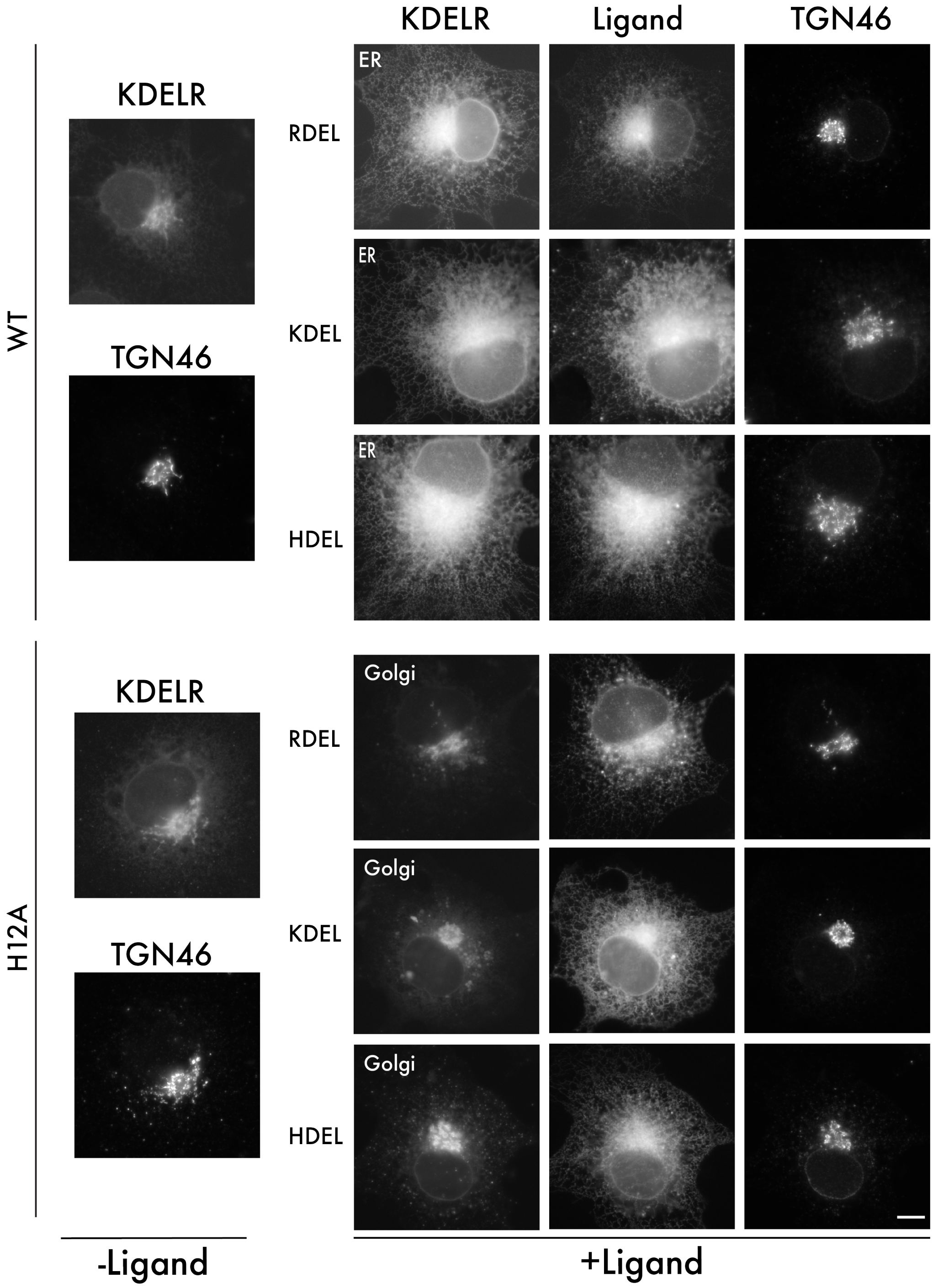


Figure 6.19 Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and an artificial K/H/RDEL ligand. KDELR mutants were tested for ligand-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

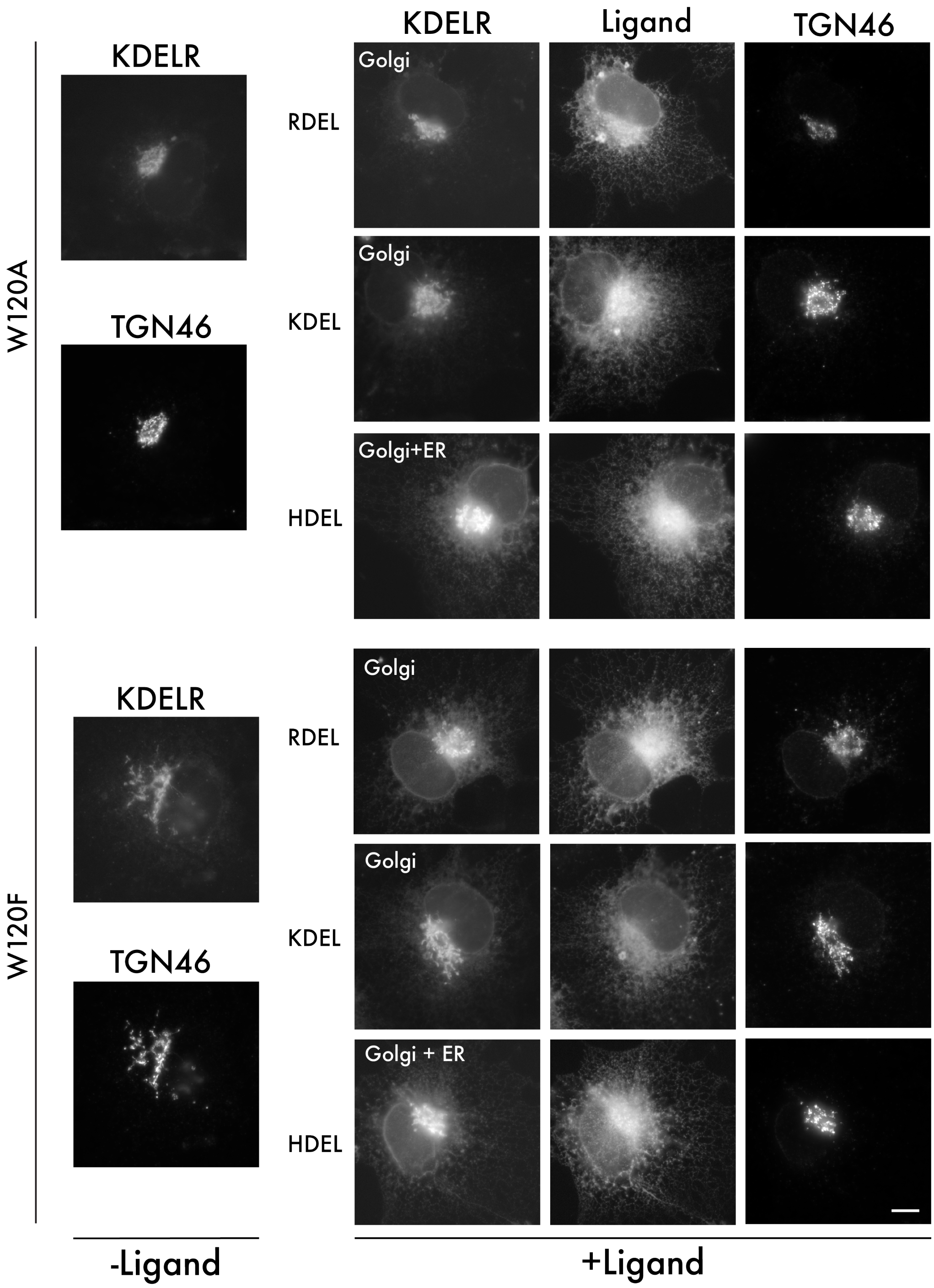


Figure 6.19 (Continued) Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and an artificial K/H/RDEL ligand. KDELR mutants were tested for ligand-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

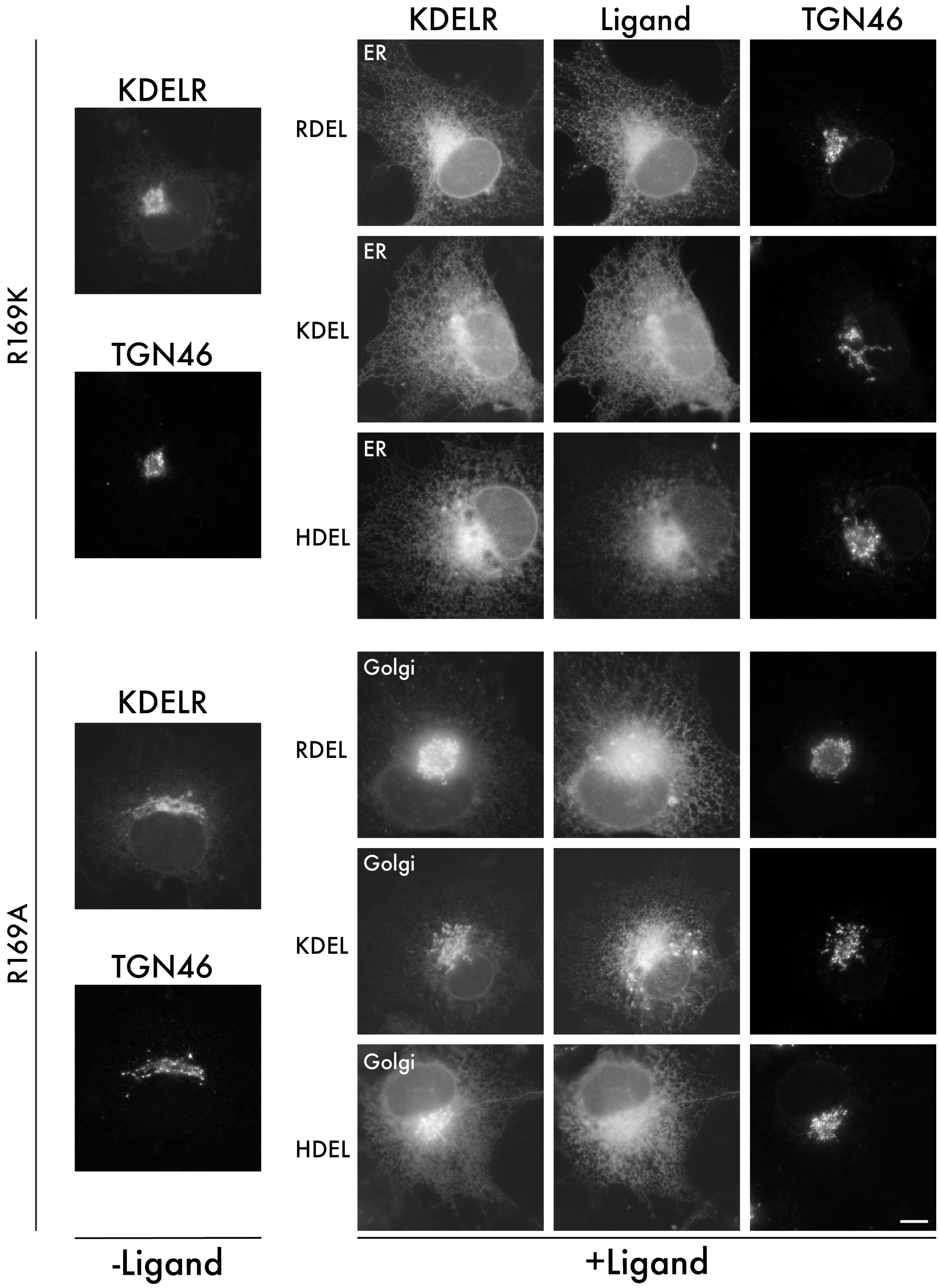


Figure 6.19 (Continued) Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and an artificial K/H/RDEL ligand. KDELR mutants were tested for ligand-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

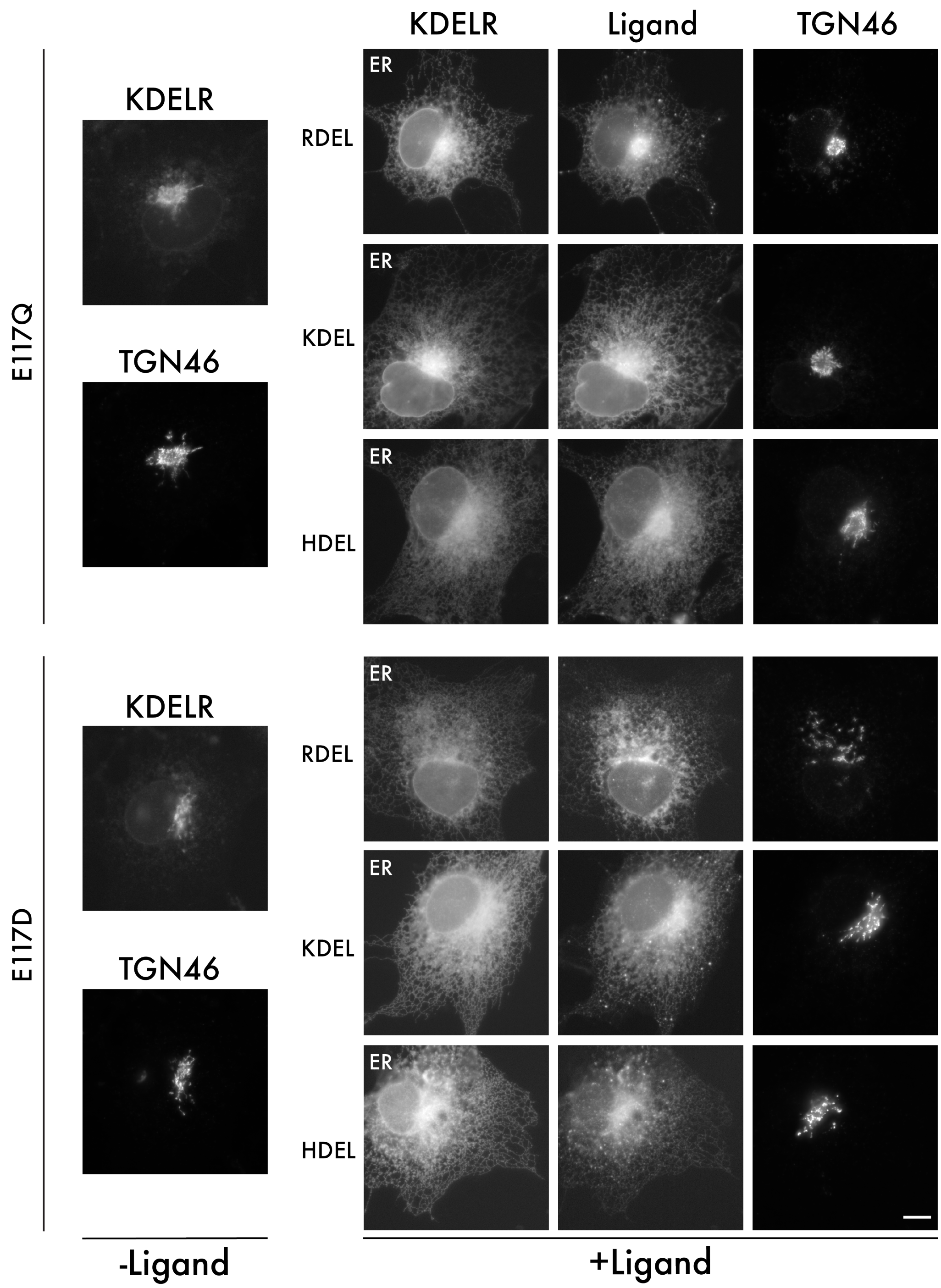


Figure 6.19 (Continued) Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and an artificial K/H/RDEL ligand. KDELR mutants were tested for ligand-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

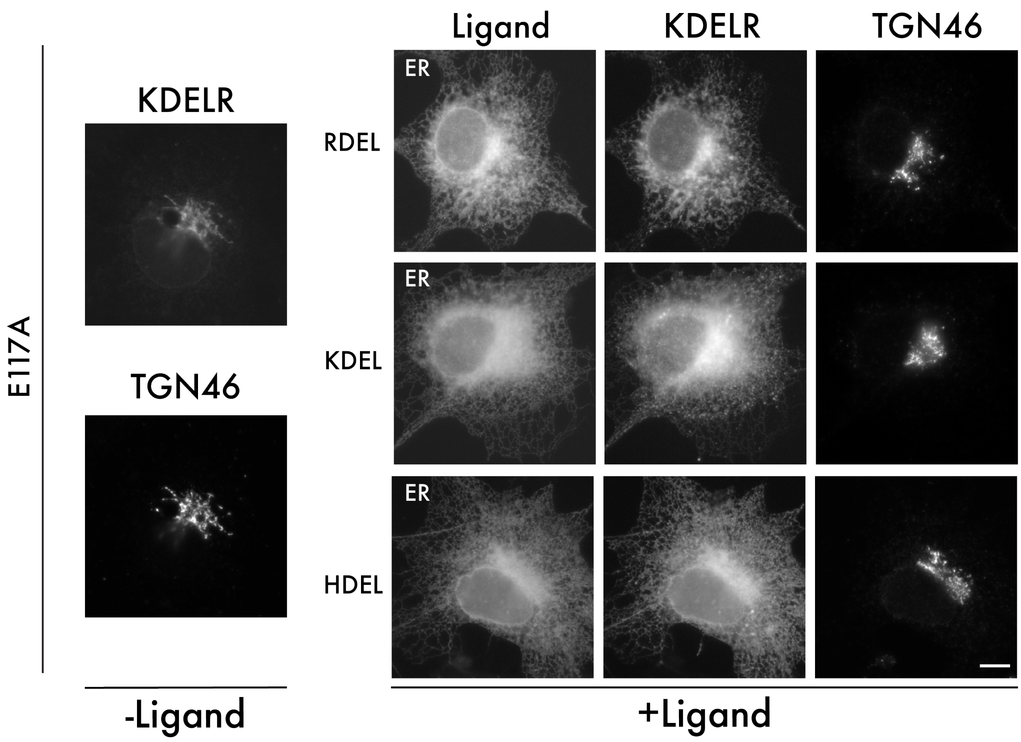


Figure 6.19 (Continued) Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and an artificial K/H/RDEL ligand. KDELR mutants were tested for ligand-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

6.6 Sybody Competition

Syb37 is observed to bind the same pocket of the KDELR as the peptide, so the KDELR-Syb37 complex should be able to effectively out-compete peptide binding, especially considering its higher affinity. In the *in vitro* assay, KDELR-Syb37 complex, prepared in the same manner as for crystallography, was incubated with ^{3}H-TAEKDEL. This complex is no longer able to bind peptide (figure 6.20), consistent with the structural and binding affinity data.

Given Syb37 is able to bind both *Mm* and *Gg* KDELR with a high affinity, it was reasoned that expressing Syb37 into the secretory pathway of COS7 cells with *Hs* KDELR1 could lead to an *in vivo* binding event, especially when considering that *Mm* and *Hs* KDELR1 differ by one amino acid. When Syb37 is co-expressed cytoplasmically (Syb37$^{\text{Cyto}}$) with KDELR-GFP, both localise as expected and do not correlate. When Syb37 is expressed in the secretory system, through the addition of a Human Growth Hormone (hGH) signal peptide (Syb37$^{\text{Sec}}$), we observed co-localisation with the KDELR and its partial redistribution from the Golgi to LAMP1-positive structures. This is consistent with the receptor no longer undergoing normal signal mediated retrieval from the Golgi, but instead flowing onwards to the lysosomes. Binding of Syb37 to the KDELR does not result in exposure of the di-lysine motif seen in the peptide-bound structure, supporting the idea that this conformational change is important in ER retrieval and trafficking.

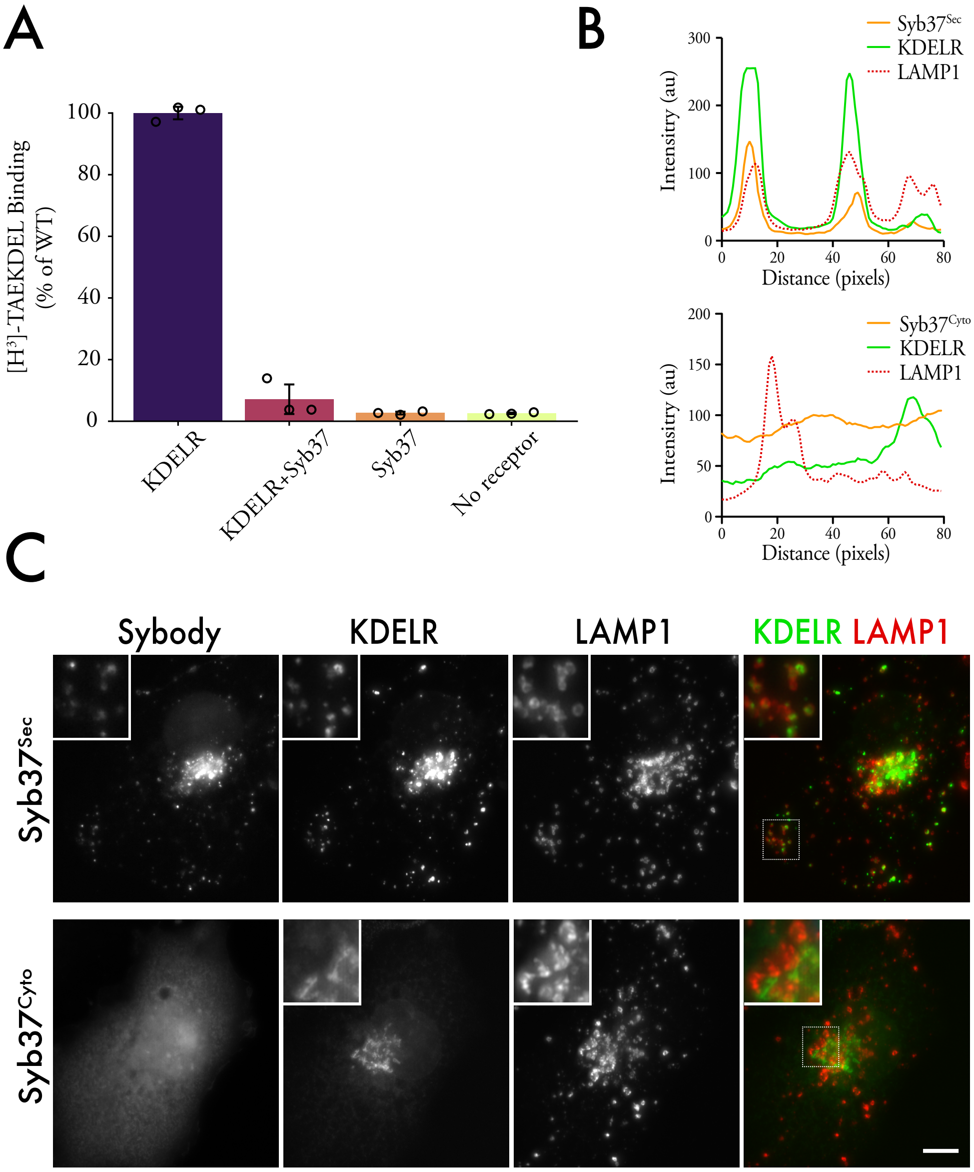


Figure 6.20 **A** Results of an *in vitro* binding assay with ^{3}H-TAEKDEL peptides to the KDELR and KDELR-Syb37 complex. **B** Line graphs of the enlarged inset regions in the images below show the co-localisation of KDELR and Syb37^{Sec} to LAMP1-positive structures. **C** Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and Syb37, once as a secreted varian (Syb37^{Sec}) and once as a cytosolic protein (Syb37^{Cyto}). KDELR mutants were tested for KDEL ligand-induced redistribution from Golgi to ER. LAMP1 was used as a lysosome marker, scale bar is 10 µm.

6.7 COPI Binding

To test potential COPI interaction sites, there was no *in vitro* assay, so the mutations in this area were probed only *in vivo*. The most obvious surface to probe is the putative lysine retrieval motif on helix 7, that is formed upon peptide binding. Townsley et al. (1993) found that mutating the 204–207 KGKK to SGSS had a small redistribution effect, but KDELR still was ER localised upon ligand stimulation. They saw no effect when mutating 206–207 KK to AA. However, in our hands, KGKK to AGAA caused ligand insensitivity, indicating a trafficking problem (figure 6.21). To benchmark this, we also made a D193N mutation, well-known to stop retrieval and also discovered by Townsley et al. (1993). This mutant also disrupted retrieval by the KDELR, while still maintaining the ability to bind KDEL peptide in the *in vitro* assay.

Studying the role of the potential COPII ER export signal is far more complex. A major issue is that a COPII binding deficient KDELR would never reach the Golgi, hence a redistribution assay becomes difficult to interpret. A further complication is that disrupting the acidic residues compromising the COPII signal play a structural role and folding defects introduced by the mutations would likely lead to ER retention. The development of an assay to probe this interaction is currently beyond the scope of this project, and is a far from trivial task.

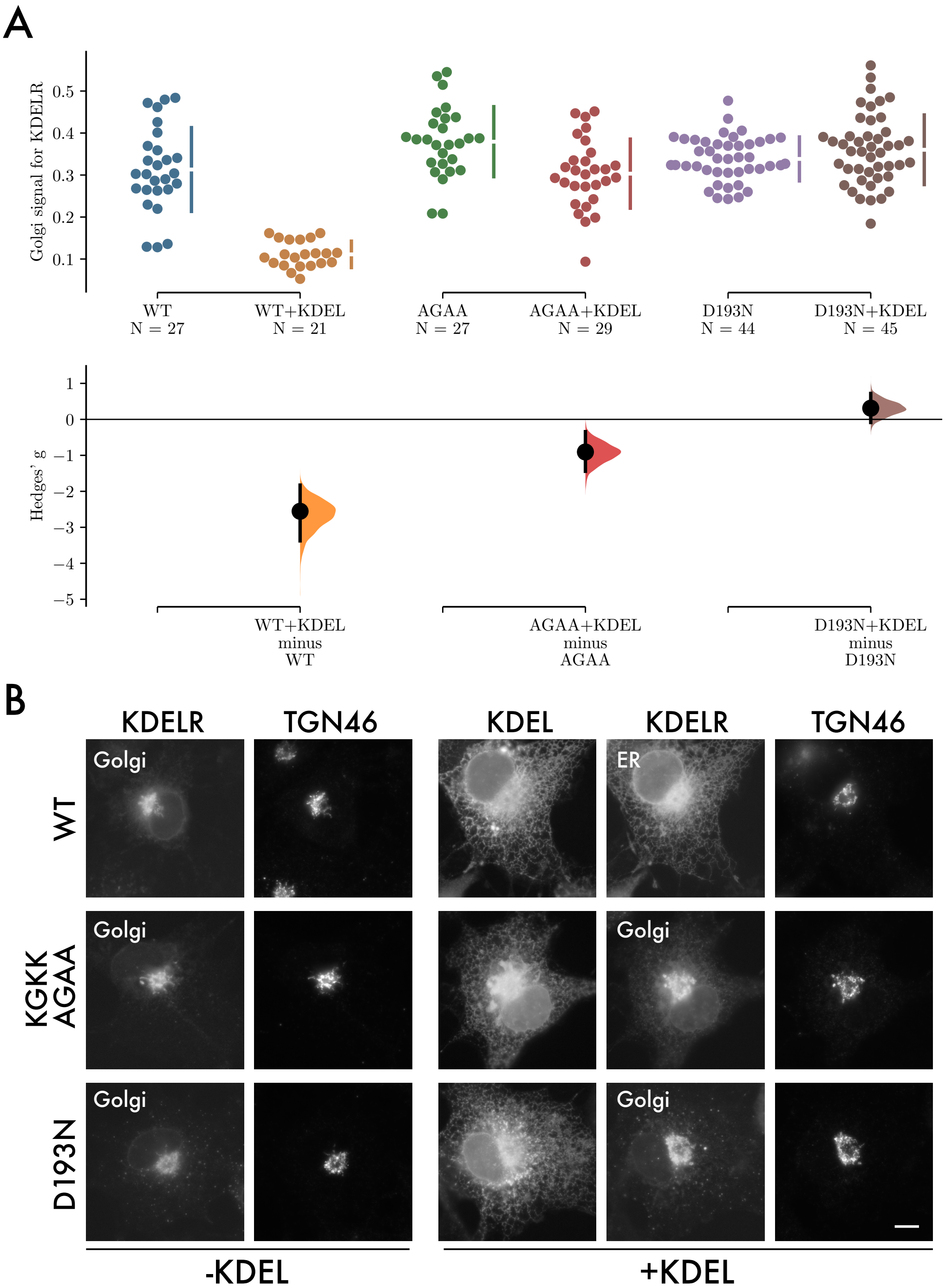


Figure 6.21 **A** Wild-type, D193N, and KGKK-dilysine motif mutant KDELRs were tested for cargo-induced redistribution from Golgi to ER. Retrieval assays were performed in the absence or presence (+) of KDEL ligand and the Golgi signal for KDELR is plotted as a Cumming plot, displaying the effect size as a bootstrap 95% confidence interval below and the standard deviation as a vertical error bar on the right of the swarm plot (Ho et al., 2019). **B** Immunofluorescence microscopy images of COS7 cells transfected with *Hs* KDELR1 and an artificial KDEL ligand (termed KDEL). KDELR mutants were tested for KDEL ligand-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar is 10 µm, all images taken at the same magnification.

6.8 Conclusion

In this chapter some of the hypotheses generated from the analysis of the structural data and study of the literature were tested by a mutational strategy in two complementary assays.

- We found evidence for the hypothesis of the short, strong hydrogen bond being involved in peptide binding and the C-terminal carboxy group of the peptide is critical for binding.
- Several residues identified in the structure are key in peptide recognition, and their relative binding to KDEL vs HDEL peptides, especially W120 mutants, corroborate structural observations.
- Syb37 binding can indeed block peptide binding and Syb37 is able to bind the KDELR *in vivo*, where it appears to lock the KDELR into the apo conformation, preventing its cycling between ER and Golgi, causing it to end up in lysosomes.
- The proposed lysine retrieval motif appears to be important in COPI retrieval, though the nature of its interaction with COPI machinery remains unclear.

There still remains much to do to improve our understanding of the KDELR's mechanisms of action. The development of more sensitive biophysical assays to interrogate binding of both peptide and other partners would be of great interest. Furthermore, the identification and *in vitro* verification of cytosolic KDELR partners would allow for a more detailed analysis of interactions between these.

7 Conclusion

At the outset of this project there were several outstanding questions regarding the KDELR. The determination of the crystal structure of the KDELR in several states, the subsequent analysis thereof and the information obtained from *in vivo* and *in vitro* assays have shed some light on how this small receptor is able to perform its critical cellular function. In brief, the questions we set out to answer were the following:

- How does the KDELR recognise the KDEL motif?
- What is the mechanism behind the pH-dependence of KDEL peptide binding?
- How does the KDELR bind COPI/COPII and how does it communicate its cargo bound state to the appropriate trafficking machinery?
- Does the KDELR share any structural similarity to transporters, and what is the role of the PQ-loop in the KDELR?

After solving the structure of the KDELR in a peptide-bound and apo state, and exploring its function with *in vivo* and *in vitro* assays, we are now able to suggest some answers to these questions. Though, as expected, these have opened up even more questions to answer in future work.

7.1 Peptide Recognition

The KDELR recognises the KDEL retention motif though a highly conserved polar cavity on its luminal side. The C-terminal peptide makes many charge-charge interactions with the receptor in a very stable conformation, allowing observation of the water networks inside the binding cavity. Steric clashes constrain the KDEL motif to the four C-terminal peptides, with the terminal carboxy group of the peptide playing a vital role. The importance of the -5 position and beyond were not explored and these residues were not well-resolved in the crystal structure.

The recognition of the H/K/RDEL variants appears to occur within the same pocket of residues, chiefly consisting of W120 and E117, with the aromatic interaction between H and W120 most likely contributing to the higher affinity of the HDEL peptide. Initial QM calculations from collaborators indicate that histidine does indeed have a more stable interaction with W120 than lysine in this context. DDEL recognition in *Kl* is likely to be mediated by the removal of negative charges with mutations of D50N, E117Q and the addition of a positive charge F175R.

The affinities of different human KDELRs to various KDEL motifs has been explored by Raykhel et al. (2007), who found that the three human homologues had differing specificities. Here we have found that in chicken, which codes for two KDELRs (75% identity to each other), *Gg* KDELR2 binds HDEL peptides with a ten-fold higher affinity to KDEL peptides, but a majority of the ER resident proteins with KDELR retention motifs contain a lysine, hinting at an additional level of control in the retention system, through different motifs.

The residues of the KDELR involved in peptide recognition are highly conserved, with the most variation permitted at the -4 and -3 positions of the motif. The basis for differential motifs is unclear both inter-species and within the paralogues expressed in the same organism. The crystal structures elucidated here provide an excellent basis for homology modelling to assist in explaining the differences observed between KDELRs.

7.2 Mechanism of Peptide Binding

The pH-dependence of KDEL peptide binding to the KDELR was confirmed with *in vitro* assays using purified, detergent solubilised KDELR. Analysis of the structure revealed a strongly conserved short hydrogen bond adjacent to the peptide binding site, flanked by a histidine residue and forming a hydrogen bonding network with the carboxy terminus of the bound peptide. This short hydrogen bond was observed to be significantly elongated in the apo structure. The residues involved in this network, especially the histidine H12, are critical in peptide binding. Furthermore, amidation of the carboxy group of the peptide severely reduces the binding affinity, indicating that this interaction is integral to peptide binding and most likely to the stabilisation of the short hydrogen bond, through the hydrogen bonding network.

Preliminary QM calculations by collaborators indicate that the stability of the hydrogen bond between E127 and Y158 depends strongly on the protonation state of H12. With fully-protonated H12 favouring sharing of the proton between the two oxygens, merging the energy wells, while deprotonation of H12 separates the energy wells.

Given the data, we suggest that the protonation state of H12 and the presence of a KDEL peptide stabilise the chemical environment around the hydrogen bond between E127 and Y158, locking the receptor in a bound state, which is destabilised upon deprotonation of H12 in response to increased pH. The order of hydrogen bond formation and the larger cytosolic conformational changes upon peptide binding is unclear, but the pH 7.0 HDEL structure indicates that peptide binding is sufficient to induce the bound-state conformation, without fully forming the short hydrogen bond.

7.3 Energetics

$$\Delta G = RT\ ln\left(\frac{[ER]}{[Golgi]}\right) \tag{7.1}$$

The question of the energy source driving the sorting by the KDELR may be somewhat of a red herring. Provided that the proton gradient between the ER and Golgi is one pH unit, or a ten-fold difference, from equation 7.1 ΔG =−5.9 kJ mol^{-1} for transport to the ER. The concentration of ER proteins has been estimated in the low mM range, but with no estimates of their concentrations in the Golgi, lets assume a µM concentration (1000:1 ratio), that gives a ΔG of 17.8 kJ mol^{-1} to overcome.

$$\Delta G = zFV_m + RT\ ln\left(\frac{[ER]}{[Golgi]}\right) \tag{7.2}$$

If an electrical potential existed between ER and Golgi, one could roughly approximate the contribution of the charges of the protons to the ΔG. The potential across the Golgi and ER membranes to the cytosol has been estimated to be roughly 0 mV. In order to arrive at the free energy to drive a thousand fold concentration gradient of ER proteins a potential of over 100 mV (from a presumed proton gradient) is required (equation 7.2).

The hydrolysis of ATP to ADP delivers ~35 kJ mol^{-1}. It must be remembered that the KDELR does not in fact transport any molecules. The KDELR binds and releases molecules upon sensing a change in environment. The transport between Golgi and ER is driven by COPI and COPII vesicle traffic, an energy dependent process. The back-of-the-envelope calculations above, when viewed in this cellular context, become irrelevant. If we treat COPI/II vesicles as constitutive and active, the KDELR merely must change affinity for its cargo under the correct conditions, which it achieves due to its pH sensitivity. We have proposed a mechanism for this, though further experiments and calculations are required to understand the pH sensitivity.

7.4 Retrieval of ER proteins by KDELR

Comparison of apo and peptide-bound structures of the KDELR revealed significant changes of the electrostatic surface on the cytosolic side. In the apo state, a negatively charged band spans the cytosolic face, reminiscent of a di-acidic COPII recognition motif. In the peptide-bound state, helix 7 kinks out, structuring and exposing several lysine residues. This disrupts the negatively charged band of the apo state and forms a new, positively charged patch. The lysines in helix 7 are similar to KK*xx* and K*x*K*xx* COPI retrieval motifs and when these are mutated to alanine, the KDELR becomes insensitive to the presence of ligand in the *in vivo* assay. Investigating and characterising the interactions that take place on the cytosolic face of the KDELR are an active area of research.

Synthesising the information obtained from structural observations and the assays performed, we propose the following model for ER protein retrieval by the KDELR, which is outlined in figure 7.1. In the Golgi the KDELR encounters an ER protein and binds the C-terminal KDEL motif. The presence of the peptide in the binding site and the protonation of H12 stabilise the formation of a short hydrogen bond between E127 and Y158. The receptor is stabilised in the bound conformation, exposing a lysine COPI retrieval motif. The KDELR and cargo protein are transported to the ER in COPI vesicles. In the ER, the increased pH leads to the deprotonation of H12, destabilising the complex and releasing the ER resident protein to the ER lumen. The KDELR adopts its apo conformation, which is recognised by COPII machinery and the KDELR returns to the Golgi, completing the cycle. In this model, the KDELR appears to couple proton transport to cargo transport using the roughly fold concentration gradient of protons between the Golgi and ER, to selectively bind cargo in the appropriate compartment. The requirement for a pH modulated lock (the hydrogen bond

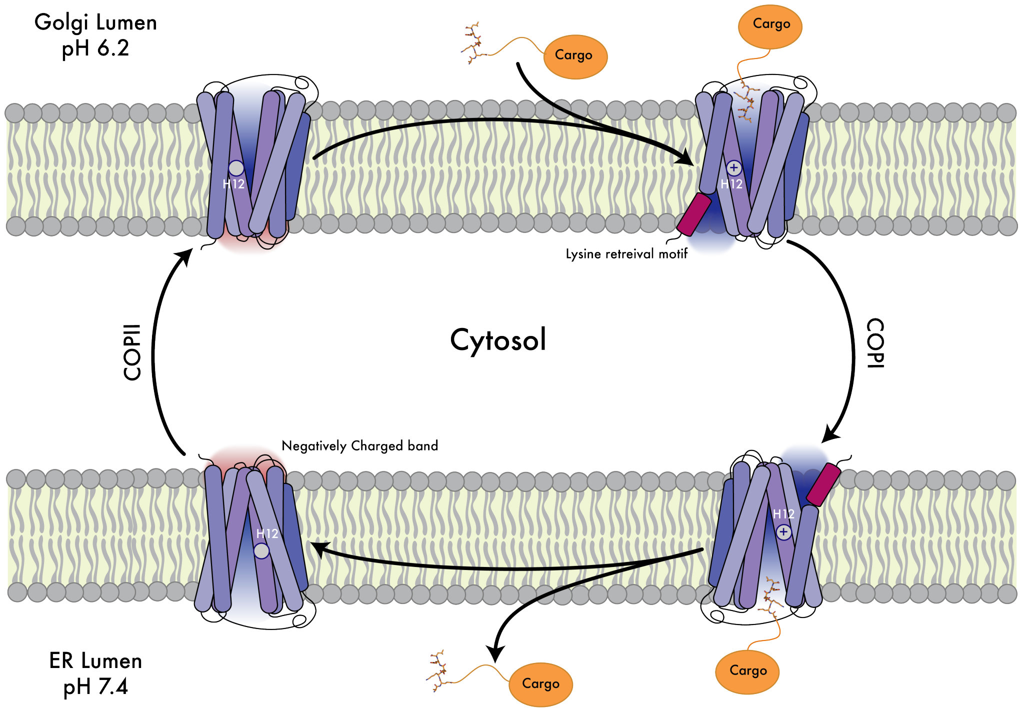


Figure 7.1 The model developed for KDELR transport is summarised in the figure above. For Golgi to ER retrieval an empty receptor must first bind to a KDEL protein in the luminal polar cavity at acidic pH. Binding results in the movement of helix 1 and 6, with the position of helix 6 being stabilised through the formation of a strong hydrogen bond interaction between Y158 on helix 6 and E127 on helix 5, which serves to lock the peptide in place. Formation of this lock is likely facilitated through the protonation of H12 on helix 1, acting as the pH sensor for the system. The movement of helix 6 causes an outward rotation of the C-terminal end of helix 7, resulting in substantial rearrangement of the electrostatic surface of the receptor facing the cytosol. The cytoplasmic negatively charged band present in the apo receptor gives way to a more basic motif on helix 7, as the C-terminal lysine side chains are exposed, revealing the likely COPI binding site. Following trafficking back to the ER, the change to neutral pH results in deprotonation of H12, causing helix 1 and 6 to move back and release the KDEL protein into the ER lumen. As helix 7 moves back, the C-terminal lysine motif packs against helix 5, re-establishing the continuous negatively charged band, which may form part of a COPII binding motif. The apo receptor is then trafficked back to the Golgi to repeat the cycle.

between E127 and Y158) that requires the presence of the KDEL peptide to form, explains how this retrieval system is able to traffic ER chaperones against their millimolar concentration gradient using the shallow pH gradient established between these two organelles. Trapping the KDELR in the apo conformation with a sybody causes it to exit the Golgi and enter lysosomes, halting the cycling between Golgi and ER, providing more evidence that the apo conformation is not recognised by COPI machinery.

7.5 Is the KDELR a Broken Transporter?

Structural homology searches with the KDELR indicate an architectural similarity to SWEET transporters. In general, the function of many PQ-loop family members, of which the KDELR is a part, have not been elucidated though it has been hypothesised that the PQ-loop family may have a dual role in transport and trafficking (**Saudek, 2017**). Structural alignments with SWEET2b, semiSWEET and PnuC show a remarkable architectural similarity, based on a THB component. It has been speculated that the THB unit forms the ancestral basis for a large group of transporters, potentially including the major facilitator superfamily (MFS) transporters (**Feng & Frommer, 2016**). It is unclear if the questions about the convergent or divergent evolution of the triple helical architecture found in different transporters will ever be answered, but in either case, the use of this structural motif points to its importance for helical transporters. Given the sequence and structural similarity of the KDELR to transporters, especially observing the cavities forming inside the KDELR, we believe the hypothesis that the KDELR derives from an ancestral transporter, possibly of peptides, is a plausible one.

7.6 Future Directions

Much work remains to be done to answer new and outstanding questions about the KDELR, some of which are listed below:

- Characterising cytosolic interactions of the KDELR, including determining affinities, oligomeric assembly and structures. Including:
  - COPI binding to peptide-bound KDELR
  - COPII binding to apo KDELR
  - Other proposed interaction partners including, G-proteins, ARF GAP1 and PKA.
- Characterising the KDELR bound to a full length luminal interaction partner, such as endogenous ER proteins and/or viral proteins and A/B toxins.
- Mutational and computational exploration of the PQ motif in the KDELR.
- Identification of sybodies that can trap the KDELR in other conformations for use in *in vivo* assays.
- Computational calculations regarding the short hydrogen bond, the carboxy terminus of the peptide and its role in the pH sensing mechanism.
- Disentangling the differential role of KDELR paralogues.
- Designing small molecule inhibitors of the KDELR as potential anti-virals.
- Determining the role of the KDELR in neurodegenerative diseases and ER stress events and how that knowledge may be utilised in treatment.

The avenues available for further research are many and will surely provide rich ground for the patient investigator. Hopefully the information presented here will provide a starting point for further enquiry and allow for the generation of further hypotheses.

Methods

8

8.1 Cloning & Mutagenesis

8.1.1 Yeast Expression Vectors

GeneArt Strings DNA Fragments of desired proteins, codon optimised for expression in *S. cerevisiae*, were ordered with 3′ and 5′ ends homologous to the pDDGFP2-Leu2d vector insertion site (Parker & Newstead, 2014). Vector was prepared by restriction digest with SmaI.

Yeast Transformation

S. cerevisiae Bj5460 was transformed in the following manner. 600 µL of an overnight culture growing in Yeast Extract–Peptone–Dextrose (YPD) was pelleted at 16,000 g for 15 s. Pellet was resuspended in 100 µL 100 mM lithium acetate, 10 mM Tris-HCl pH 7.4. To this a mixture of 10 µL ssDNA mixed with 2 µL of 150 ng μL^{-1} pDDGFP2-Leu2d digested with SmaI and 10 µL of 200 ng μL^{-1} DNA fragment was added. Subsequently 400 µL of 100 mM lithium acetate, 10 mM Tris-HCl pH 7.4 with 50% (w/w) PEG 3350 was added and the mixture was incubated at room temperature on a rotating wheel for 15 min. 50 µL of DMSO was added and the cells were incubated at 42 °C for 15 min. Cells were then pelleted by centrifugation at 500 g for 30 s, resuspended in 75 µL water and spread onto a selective minus Uracil dropout media plate for 3 days at 30 °C, after which they were plated on a minus Leucine dropout media plate for another 3 days.

Yeast Miniprep

5 mL of transformed yeast was grown overnight at 30 °C in minus Leucine dropout media. Cells were harvested by centrifugation at 3500 g for 5 min and resuspended in 200 µL PBS. 150 µL silica beads were added and the mixture was vortexted for 10 min at 4 °C. The resulting lysate was used in a standard miniprep kit according to the manufacturer's instructions.

Subsequent transformations of BJ5460 were then carried out with the pDDGFP-Leu2D vector now containing the desired protein and thus only vector was added to the ssDNA mixture.

8.1.2 Mammalian Expression Vectors

Codon optimised *Hs* KDELR1 was purchased in a pMK vector and subcloned into a pEF5/TO/FRT vector with a C-terminal linker (five copies of Gly-Ser-Ser-Ser (GS3)5) followed by eGFP. Codon optimised Sybody37 was cloned into a pcDNA3.1+ vector with a C-terminal (GS3)5 linker followed by mScarlet. Another Sybody37 construct had an additional 26 residue N-terminal hGH secretion signal peptide. Similarly an mScarlet with an N-terminal hGH sequence and the 16 C-terminal residues of *Hs* BiP at its C-terminus was cloned into a pcDNA4 vector.

8.1.3 Sybody Cloning

Sybodies from the selection process were subcloned from the pNb_init vector into an MBP fusion expression vector pBXPHM3 or pBXPC3H via the Fragment Exchange (FX) cloning strategy using BspQI (Geertsma, 2014). Briefly target vector and the vector containing the gene fragment of interest are combined with BspQI, which is then heat deactivated after sufficient time for digestion. Subsequently a T4 ligase is added to the mixture with ATP and then transformed into chemically competent *E. coli* and plated onto LB agar plates with appropriate antibiotic resistance. The target vector contains a ccdB suicide cassette, thus only vector with the desired fragment will form colonies on the agar plate.

8.1.4 Mutagenesis

Mutations were introduced by site directed mutagenesis in the pMK vector. Complimentary forward and reverse PCR primers were designed containing the mutation desired. PCR was carried out and subsequently treated with the DpnI restriction enzyme. The digested PCR reaction was then transformed into One Shot® OmniMAX™ 2 T1^R *E. coli* and plated onto LB agar plates with appropriate antibiotics. Mutated sequences were then subcloned from pMK into the desired vectors (pEF5/TO/FRT or pDDGFP-Leu2D).

8.2 Protein Expression

8.2.1 *S. cerevisiae* Expression

S. cerevisiae BJ5460 was transformed and plated onto minus Uracil drop out agar plates for three days, then moved to -Leucine drop out agar plates for another three days. Then yeast was grown in 100 mL overnight cultures in minus Leucine drop out media containing 2% glucose, after which it was moved to a 1 L culture in the same media, grown overnight. The large overnight culture was then either split into 12 flasks containing 800 mL minus Leucine drop out media with 2% lactate or into a fermenter with 13 L of minus Leucine drop out media with 2% lactate. Flasks were induced with 2% galactose after 18 h and harvested 26 h after induction, cells were resuspended in PBS after centrifugation. In the fermenter, induction occurred after 18 h and harvest generally 24 h after induction, cells were resuspended in PBS after centrifugation. Yeast was grown at 30 °C at 240 rpm in flasks. For fermenter growth, pH was held constant at 5.1.

8.2.2 Sybody Expression

E. coli MC1061 were transformed with the desired Sybody in the pBXPMH3 or pBXPC3H vector and plated onto LB agar with 120 $\mu g mL^{-1}$ ampicillin. One colony was used to start an overnight culture in TB media with 100 $\mu g mL^{-1}$ ampicillin. The next day six flasks containing 1 L TB were inoculated with the overnight culture. Cells were grown at 37 °C shaking at 200 rpm. At an OD_{600} of 0.5 the culture was cooled to 25 °C and expression was induced overnight by the addition of 0.02% L-arabinose. Cells were harvested by centrifugation.

8.3 Membrane Fraction Preparation

Cells were resuspended in PBS and lysed in a cell disruptor (Constant Systems) at 38 kpsi. Subsequently lysate was centrifuged at 27,500 g for 30 min. Supernatant was then centrifuged at 235,000 g for 1.5 h. Pellet was resuspended in 1 M potassium acetate, 20 mM HEPES pH 7.5 and homogenised in a Dounce homogeniser. A final centrifugation step of 235,000 g for 1 h was carried out and the pellet was resuspended in PBS.

8.4 Protein Purification

8.4.1 Fluorescence Size Exclusion Chromatography

Membrane fractions were solubilised in 1% (w/v) detergent for 1 h at 4 °C and then centrifuged at 30,000 g for 30 min, fluorescence was measured before and after to check efficiency. The supernatant was then run on a Superose 6 SEC column in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.03% DDM, 0.006% CHS, detecting GFP fluorescence (488→512 nm) at a flow rate of 0.4 mL min^{-1}.

8.4.2 KDELR Purification

Membrane fractions were solubilised in 1% (w/v) detergent (1% DDM - 0.05% CHS), PBS supplemented with 150 mM NaCl, 10% glycerol for 2 h then centrifuged at 235,000 g for 1 h. Supernatant was allowed to bind Ni^{2+} resin (Thermo Scientific HisPur^{TM}) for 5 h or until binding plateaued with the addition of 15 mM imidazole. Resin was washed in a gravity column with 0.1% detergent, PBS supplemented with 150 mM NaCl, 10% glycerol and 15 mM imidazole, followed by a second wash with 38 mM imidazole. Protein was eluted with 250 mM imidazole. 5 mg of TEV protease (purified in house) was added to the eluant and left to dialyse overnight at 4 °C in 2.5 L 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 3x the Critical Micelle Concentration (CMC) of the detergent in use, in a 3.5 kDa cut off regenerated cellulose membrane. Dialysed protein then underwent a reverse IMAC, in dialysis buffer, where flow through was collected. Flow through was concentrated in a 50,000 Mw concentrator (Amicon® Ultra, ultracel regenerated cellulose) to 0.5 mL and loaded on an S200 SEC column in 20 mM Tris-HCl pH 7.5, 100 mM NaCl. Fractions containing protein were run on SDS-PAGE, pooled and concentrated in a 50,000 Mw concentrator (Amicon® Ultra) to desired concentration.

8.4.3 Sybody Purification

Cell pellets were resuspended in 20 mM Tris-HCl pH 7.5, 150 mM NaCl and lysed in a cell disruptor (constant systems) at 30 kpsi. Lysate was centrifuged at 8000 g for 20 min. 15 mM imidazole was added to the supernatant and this was incubated with 10 mL Ni-NTA beads (Thermo Scientific HisPur^{TM}) for 1 h. The beads were then loaded onto a gravity flow column and washed with 100 mL lysis buffer with 50 mM imidazole. Protein was eluted with lysis buffer supplemented with 300 mM imidazole. Protein was then dialysed overnight at 4 °C against 3 L 20 mM Tris-HCl pH 7.5, 150 mM NaCl with the addition of 500 µg 3C protease (purified in house). After dialysis, protein was run over a 5 mL HisTrap HP column in dialysis buffer with 15 mM imidazole, the flow through was collected and concentrated. SEC was performed on an SRT-C SEC 300 (Sepax) column in 20 mM Tris-HCl pH 75, 150 mM NaCl. Desired fractions were pooled and concentrated with an Amicon® Ultra concentrator with a molecular weight cut-off of 3 kDa to desired concentration.

8.4.4 Circular Dichroism

CD measurements were performed on a Jasco J-815 Spectropolarimeter. Buffers were 10 mM potassium phosphate, buffered to the desired pH, 50 mM Na_2SO_4 and 3x CMC of the detergent used for protein purification. Peptide was added at 20 µM.

8.4.5 nanoDSF

NanoDSF measurements were performed with a NanoTemper Prometheus, in standard capillaries. Protein was added at 5 µM and peptide at 20 µM, in the stated buffers. A 20–90 °C temperature gradient was run at 2 °C min^{-1}.

8.4.6 SEC-MALS

SEC-MALS experiments were performed in the molecular biophysics facility by Dr. David Staunton, using an S200 increase SEC column, Wyatt Heleos8+ and T-rEX detectors and analysed with Wyatt ASTRA software.

8.5 Sybody Generation

8.5.1 Biotinylation

Target KDELR had an Avi tag `GLNDIFEAQKIEWHE` inserted before the TEV protease site in the pDDGFP Leu2d vector. Protein was expressed and purified as described above. Protein sample was then concentrated to about 40 µM and biotinylated using the Avidity BirA biotin-protein ligase reaction kit according to the manufacturer's instructions. In brief, BirA biotinylates the Avi-sequence via an biotinyl 5'-adenylate intermediate on the lysine in the Avi sequence. This reaction is carried out overnight at 4 °C after which the protein is loaded onto a SEC column in the same manner as during purification.

To test for biotinylation efficiency a streptavidin gel shift assay is performed. Streptavidin is added, in 2–5x molar excess, to biotinylated sample in NuPAGE™ LDS sample buffer and run on an SDS-PAGE at 120 V on ice.

8.5.2 Ribosome Display

10 µL of PUREfrex®SS translation mix was prepared, 1 µL of 2 µM library mRNA was added and the mix was incubated at 37 °C for 30 min. Then the ribosomal complexes were diluted in 100 µL buffer (either 20 mM Tris-HCl pH 7.5 or 50 mM potassium phosphate pH 5.0, 100 mM NaCl, 50 mM $MgCl_2$, 0.1% DDM, 0.005% CHS, 0.5% BSA, 5 mg mL^{-1} Heparin and 40 U of RNasin). The reaction was then centrifuged at 21,000 g for 5 min at 4 °C. Then 50 nM of biotinylated target protein was added to the panning mix, as well as 200 nM TAEHDEL peptide in the case of the pH 5.0 samples. This was incubated for 20 min on ice. 12 µL Dynabeads MyOne Streptavidin T1 were washed twice with 500 µL Buffer A (20 mM Tris-HCl pH 7.5, 100 mM NaCl) or Buffer B (50 mM potassium phosphate pH 5.0, 100 mM NaCl) with 0.5% BSA, 0.1% DDM and 0.005% CHS and then blocked in the same solution for 30 min. Panning mix was added to the beads and incubated for 5 min. Beads were then washed three times with 500 µL Buffer A or B with 0.03% DDM, 0.0015% CHS and 200 nM TAEHDEL peptide in the case of buffer B, during the last wash step beads were placed into a fresh tube. RNA was eluted from the beads with 100 µL 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 50 mM EDTA pH 8.0 and 100 µg mL^{-1} yeast RNA and incubated at room temperature for 10 min. The eluted RNA was purified using the RNeasy micro kit from Qiagen using the manufacturer's instructions and eluted into 14 µL RNAse free water. Reverse transcription was performed on the eluted

RNA by adding 2 µL of the RT_Primer at 100 µM and 4 µL of 10 mM dNTPs and incubating at 65 °C for 5 min and then cooling on ice. To this mixture a 40 µL reverse transcription reaction was set up according to the manufacturer's instruction (Affinityscript, Agilent) and incubated for 1 h at 37 °C followed by 5 min at 95 °C. cDNA was purified using a PCR purification kit (Macherey Nagel) and eluted in 30 µL elution buffer. The cDNA was used in a PCR using Med_FX_for/ Med_FX_rev for concave/loop sybodies or Long_FX_for/ Long_FX_rev for convex sybodies to add FX overhangs to the DNA. Amplified sybody pools and pDX_init were digested with BspQI, gel-purified and 500 ng sybody insert and 1 µg pDX_init backbone were ligated in 50 µL using 5 U of T4 DNA Ligase. A qPCR step was also performed on sybody cDNA to quantify full length cDNA and total amount of cDNA present.

8.5.3 Phage Display

The sybody pools ligated with pDX_init were transformed in 350 µL electrocompetent *E. coli* SS320. After electroporation (BioRad Gene Pulser, 25 kV, 200 Ω, 25 pF) cells were resuspended in 25 mL SOC media and shaken at 37 °C for 30 min. Cells were diluted into 250 mL of 2YT media with 2% glucose, 100 µg mL^{-1} ampicillin and grown overnight, shaking at 37 °C. Of this overnight culture 1 mL was diluted into 50 mL 2YT, 2% glucose, 100 µg mL^{-1} ampicillin and grown to an OD_{600} of 0.5. 10 mL of culture was infected with $3 \cdot 10^{11}$ plaque forming units of M13KO7 helper phage at 37 °C without shaking for 30 min. Infected cells were centrifuged and resuspended in 50 mL 2YT, 100 µg mL^{-1} ampicillin, 25 µg mL^{-1} kanamycin and incubated at 37 °C, shaking overnight. Cells were pelleted and 40 mL of the culture supernatant was mixed with 10 mL 20% PEG 6000 (v/v), 2.5 M NaCl and incubated on ice for 30 min. Precipitated phages were pelleted by centrifugation. The pelleted phages were resuspended in 1 mL PBS and centrifuged at full speed on a table top centrifuge. This clearing procedure was carried out twice. Phage concentration was determined by UV-vis spectroscopy at 269 nm and 320 nm (Phage·mL^{-1} = $(A_{269}-A_{320})\cdot 6\cdot 10^{16}/4900$).

A Maxi-sorp plate was coated with 100 µL per well of 60 nM neutravidin in buffer A or B overnight. The plate was then washed three times with buffer A or B and blocked with buffer A or B containing 0.5% BSA for 1 h. Phages were diluted to 10^{12} phages mL^{-1} in Buffer A or B containing 0.5% BSA, 0.1% DDM and 0.005% CHS and incubated with 50 nM biotinylated target protein, with 200 nM of TAEHDEL in buffer B, for 20 min. The plate was washed three times with Buffer A or B containing 0.5% BSA, 0.1% DDM and 0.005% CHS and 100 µL of phage-target protein mixture was added to each well and allowed to incubate for 10 min at 4 °C. The plate was then washed three times with Buffer A or B with 0.03% DDM and 0.0015% CHS and 200 nM of TAEHDEL peptide in buffer B. Elution of phages was carried out by adding 100 µL of buffer A with 0.25 mg mL^{-1} trypsin for 30 min at room temperature. 0.125 mg mL^{-1} 4-(2-Aminoethyl)benzenesolfonyl fluoride was added to inhibit trypsin. 5 mL of eluted phages were added to 45 mL of *e. coli* SS320 in 2YT with an OD_{600} of 0.5 and incubated at 37 °C without shaking for 30 min.

The cells were then diluted into 250 mL 2YT, 2 % glucose, 100 µg mL^{-1} ampicillin and grown overnight at 37 °C resulting in a preculture to produce more phage for the second round. The second round of phage display was performed in the same manner as above, except that magnetic beads (Dynabeads MyOne Streptavidin C1) were used. The resulting culture of infected cells was then used to purify the phagemids (MiniPrep, Qiagen) containing the selected sybodies. Sybody sequences were subcloned into pNb_init using FX cloning and transformed into *E. coli* MC1061 for ELISA analysis and protein purification.

8.5.4 ELISA

Single sybody clones were picked and expressed in 1 mL TB with 25 µg mL^{-1} chloramphenicol in a 2 mL 96 deep well plate. Expression was induced with 0.02% arabinose at 22 °C overnight. After expression, the cells were pelleted and then resuspended in 50 µL B-PER II for lysis. Lysate was diluted with 950 µL buffer A and centrifuged to pellet cell debris. ELISAs were carried out in Maxi-

Sorp plates coated with 100 µL per well of 5 µg mL^{-1} protein A in buffer A. The plate was washed 3 times with 250 µL buffer A and blocked with 250 µL buffer A with 0.5% BSA. Plates were then washed three times with buffer A and incubated with the c-myc antibody (1:2000, Sigma 9E10) diluted in buffer A with 0.5% BSA, 0.1% DDM, 0.005% CHS for 1 h. Plates were then washed three times with buffer A with 0.03% DDM, 0.0015% CHS. The plates were then incubated by crude lysate diluted five-fold with buffer A, 0.5% BSA, 0.1% DDM, 0.005% CHS and incubated for 1 h. Plates were washed three times with either buffer A or B (as appropriate for the selected sybody). 50 nM of biotinylated target protein was added to the wells in buffer A or B with 0.1% DDM, 0.005% CHS as well as 200 nM TAEHDEL peptide for buffer B, and incubated for 20 min. Plates were then washed three times in Buffer A or B with 0.03% DDM, 0.0015% CHS and incubated with Streptavidin-HRP (1:5000) in buffer A or B with 0.5% BSA, 0.1% DDM, 0.005% CHS for 10 min. After three final washes in the corresponding buffers the ELISA was developed by adding 100 µL of 0.1 mg mL^{-1} TMB, 50 mM Na_2HPO_4, 25 mM Citric acid, 0.006% H_2O_2. Results were measured at an absorbance of 650 nm.

8.5.5 Small Scale Purification

Sybodies selected from the ELISA were transformed into *E. coli* MC1061. 100 µL of the transformation recovery mix was used to inoculate a 1.2 mL overnight culture in TB with 25 µg mL^{-1} chloramphenicol at 37 °C. 50 mL TB cultures with 25 µg mL^{-1} chloramphenicol were inoculated with 840 µL from the overnight and allowed to grow for 1.5 h at 37 °C. The cultures were then cooled to 22 °C for one hour, induced with 0.02% arabinose and left over night. Cells were pelleted and resuspended in 5 mL osmotic lysis buffer, 50 mM Tris-HCl pH 8.0, 20% sucrose, 0.5 mM EDTA, 5 µg mL^{-1} lysozyme. The mixture was allowed to rest for 30 min at 4 °C, then 20 mL buffer A was added with 1 mM $MgCl_2$ and lysate was centrifuged at 4000 g for 15 min. Supernatant was decanted and incubated with 800 µL Ni-NTA resin for 1 h, after which the mixture was centrifuged and the supernatant discarded. Beads were washed twice with 500 µL buffer A with 30 mM imidazole on at His MultiTrap^{TM} HP plate. Sybodies were eluted in 550 µl buffer A with 300 mM imidazole. 500 µL of the eluate was injected onto a SRT-10C SEC 300 column, the desired fractions were pooled and used for additional ELISA experiments and subsequent SPR experiments.

8.5.6 SPR Screening

SPR experiments were carried out on a Biacore T200 with a Biotin CAPture sensor chip. Experiments were performed at 20 °C in 20 mM Tris-HCl pH 7.5, 150 mM NaCl with 0.03% DDM, 0.0015% CHS.

CAPture reagent was loaded onto the sensor chip according to the manufacturer's instructions. For Sybody screening, chips were loaded with 2 µM biotinylated KDELR for 300 s at 2 µL min^{-1} followed by a 180 s wash with buffer at 30 µL min^{-1}. Affinity measurements began with a 60 s wash step, followed by 180 s of binding, with Sybodies concentrated at 1 µM. Followed by a 600 s dissociation step in buffer. After each sybody run, the chip was stripped with regeneration solution (250 mM NaOH, 6 M guanidine-HCl).

All experiments were performed with 2-1 flow path, subtracting the reference measurement of the path with no biotinylated KDELR attached. Reference measurements with buffer only were also taken and subtracted from the data.

8.5.7 Bio-layer Interferometry

BLI was performed on an Octet Red 384 system using streptavidin biosensors (SA). Biosensors were equilibrated in buffer (20 mM Tris-HCl, 150 mM NaCl, 0.03% DDM, 0.0015% CHS) for 15 min. Avi-tagged KDELR was then immobilised on the sensors with a 60 s baseline step in buffer, followed by a 300 s association step in Avi-tagged KDELR at 0.1 mg mL^{-1} and another 60 s baseline step in buffer. Kinetics were measured with a serial dilution of the Sybody, with one reference well. There was a 60 s baseline step in buffer, followed by a 300 s association step in the desired sybody concentration, followed by a 300 s dissociation step in buffer. The experiment was repeated with reference sensors that had not seen Avi-tagged KDELR. Data was analysed in the Octet v9.0 software package. All raw data was baseline and reference subtracted, in-step corrected, y-axis aligned and filtered with a Savitzky-Golay filter.

8.5.8 qPCR

qPCR was performed with a 7500 fast qPCR machine. Reactions contained SYBR select Master Mix, 300 nM of each primer, 5% DMSO and 2 µL cDNA from ribosome display or phages (diluted 10:1 in water). Standard curves for relevant primer pairs had been previously determined (**Zimmermann et al., 2018a**).

8.6 Crystallography

8.6.1 Lipidic Cubic Phase

KDELR Apo

Gg KDELR was concentrated to 15.5 mg mL^{-1} and a protein containing lipid mesophase with MO was prepared by homogenisation in a dual syringe mixing device at 20 °C, in a 60:40 (w/w) ratio. Crystallisation was performed at 4 °C in a 96 well glass sandwich plate using 50 nL mesophase and 0.8 µL precipitant solution. Crystals were harvested after 10 days in the precipitant condition consisting of 30% (v/v) PEG 500 DME, 100 mM Tris pH 9.0 and 100 mM magnesium sulphate. Wells were opened using a tungsten carbide glass cutter, crystals harvested using 30–75 µm micromounts and cryocooled in liquid nitrogen.

KDELR TAEKDEL Bound

Gg KDELR was concentrated to 14.5 mg mL^{-1} and incubated with 6.4 mM TAEKDEL peptide (from a 100 mM stock solution in water) on ice for one hour. Crystals were set up and harvested as above but crystals formed at 20 °C in the precipitant 30% (v/v) PEG 600, 100 mM MES pH 6.0, 100 mM sodium nitrate and were harvested as above.

KDELR TAEHDEL Bound pH 6.0

Gg KDELR was concentrated to 14.5 mg mL^{-1} and TAEHDEL peptide was added as above. Crystals formed at 20 °C in the precipitant 30% (v/v) PEG 550 MME, 100 mM MES pH 6.0, 100 mM ammonium phosphate dibasic and were harvested as above.

KDELR TAEHDEL Bound pH 7.0

Gg KDELR and TAEHDEL peptide were prepared as above, and were set up from the same LCP batch as the pH 6.0 TAEHDEL crystals. Crystals formed at 20 °C in the precipitant 30% (v/v) PEG 500 DME 100 mM HEPES pH 7.0, 100 mM Potassium thiocyanate and were harvested as above.

Mm KDELR Sybody Bound

Mm KDELR Syb37 complex was concentrated to 15.84 mg mL^{-1}, after being bound on ice for 1 h and run on an SRT-C SEC 300 column (Sepax). Crystals were set up as above and formed at 20 °C, 30% (v/v) PEG 500 DME, 100 mM Tris pH 9.0, 100 mM sodium nitrate and were harvested as above.

Gg KDELR Syb37 Bound

Gg KDELR Syb37 complex was concentrated to 15.4 mg mL^{-1} after being bound on ice for 1 h and run on an SRT-C SEC 300 column (Sepax). Crystals were set up as above and formed at 20 °C in the precipitant 30% (v/v) PEG 400, 100 mM Tris pH 9.0 and were harvested as above.

8.6.2 Data Collection & Processing

All data except the *Mm* KDELR-Syb37 was collected on the I24 microfocus beamline at Diamond Light Source. Indexing, scaling and merging was performed using the Xia2 pipeline to DIALS and AIMLESS or to XDS and XSCALE (**Winter, 2010**; **Winter et al., 2018**; **Evans & Murshudov, 2013**; **Kabsch, 2010**). *Mm* KDELR-Syb37 data was collected at the Proxima2A beamline at the Soleil Synchrotron, data was indexed and integrated with XDS, scaled and merged in AIMLESS (**Kabsch, 2010**; **Evans & Murshudov, 2013**; **Evans, 2011**).

Intial phases were determined by molecular replacement of the KDELR-Syb37 complex using Phaser with PDB:5M13 as a search model (**McCoy et al., 2007**). Using the map from these initial phases a model was built into the density using O and Coot (**Jones et al., 1991**; **Emsley et al., 2010**). The model was then refined in Phenix and BUSTER (**Adams et al., 2010**; **Bricogne et al., 2017**). Apo and peptide-bound structures were then phased using molecular replacement with the KDELR model from the KDELR-Syb37 complex crystal structure. The distance of the short hydrogen bond in the peptide-bound complex (PDB: 6I6H) was fixed using the LINK command in BUSTER. However, refinement in Phenix suggested a shorter distance of ~2.3 Å.

8.6.3 Experimental Phasing

Fluorescence Detection of Heavy Atom Labelling

Heavy atom solution is added to 10 µg of protein in 1 or 3 fold molar excess and incubated for 1 h at room temperature. Then ten-fold excess of TMRM is added and incubated for another 15 min. Samples are run on SDS-PAGE and first imaged with UV light, followed by staining with coomassie brilliant blue.

Heavy Atom Screening

Heavy atom solution is added to 60 µL of protein at 20 µM in large molar excess and incubated for 30 min on ice. Sample is centrifuged for 10 min at 30,000 g. Supernatant is loaded onto an S200 10:50 column.

Heavy Atom Soaking

Cover glass of the LCP plates was cut open using a tungsten carbide glass-cutter and was removed. 1 µL solution from the screen, supplemented with the desired concentration of heavy atom compound, was added to the crystal drop and the plate resealed. Crystals were then harvested as before.

8.7 Binding Assays

8.7.1 ^{3}H-TAE[KH]DEL Binding

Assays were performed in 20 mM MES pH 5.4, 40 mM NaCl, 0.01% DDM 0.0005% CHS unless stated otherwise. 5 µL of ^{3}H-TAEKDEL or TAEHDEL at 20 nM was added to 5 µL of *Gg* KDELR or mutants thereof at the desired concentration and incubated at room temperature for 12 min. For competition assays with unlabelled peptides, a dilution series of unlabelled peptide was made over the desired range at a final volume of 4 µL. Then 5 µL of *Gg* KDELR was added at the desired concentration, followed immediately afterwards with 1 µL of labelled peptide, with at most a 1 min delay due to pipetting. The mixture was then bound to 0.22 µm mixed cellulose ester filters. Filters were then washed with 2 x 500 µL buffer. The amount of peptide remaining bound was measured by scintillation counting in Ultima Gold (Perkin Elmer). Specific activity was 106 Ci mmol^{-1} for ^{3}H-TAEHDEL and 128 Ci mmol^{-1} for ^{3}H-TAEKDEL. Experiments were performed at least three times. Results were fit with a three parameter sigmoid $y = \frac{\max \cdot x^h}{K_d^h + x^h}$ using scipy, while competition assays were fit to four parameters $y = \frac{\min + (\max - \min)}{1 + \frac{K_d^h}{x^h}}$, with min and max bounded to 0, and 1.

8.8 Immunofluorescence Microscopy

8.8.1 COS7 Cell Culture

COS7 cells were grown in DMEM containing 10% bovine calf serum at 37 °C and 5% CO_2. Cells were plated at 50,000 cells per well of a 6-well plate, on-top of two glass coverslips with 10 mm diameter and a thickness of 0.16–0.19 mm. For ER retrieval assays, cells were transfected after 18 h with 0.5 µg *Hs* KDELR-GFP (pEF5/TO/FRT) and 0.5 µg mScarlet-KDEL ligand (pcDNA4) or 0.5 µg *Hs* KDELR-GFP (pEF5/TO/FRT) and 0.5 µg pcDNA4 (no ligand), diluted in 100 µL Optimem and 3 µL Mirus LT1. For Golgi localisation assays with Syb37 cells were transfected with 0.5 µg KDELR-GFP (pEF5/TO/FRT) and 0.5 µg Syb37^{Sec} (pcDNA3.1+) or 0.5 µg KDELR-GFP (pEF5/TO/FRT) and 0.5 µL Syb37^{Cyto} (pcDNA3.1+).

8.8.2 PLP Staining

Cells were grown on glass cover slips and 22 h after transfection were washed twice with 2 mL of PBS, then fixed for 2 h in 2 mL PLP (2% (w/v) paraformaldehyde in 87.5 mM lysine, 87.5 mM sodium phosphate pH 7.4, and 10 mM sodium periodate). Subsequently cover slips were washed three times in 2 mL permeabilisation solution (100 mM sodium phosphate pH 7.4), then permeabilised in 1 mg mL^{-1} BSA, 0.12 mg mL^{-1} saponin, and 100 mM sodium phosphate pH 7.4 for 30 min. Primary and secondary antibody staining was performed for 1 h in permeabilisation solution at room temperature. Commercially available antibodies were used to TGN46 (sheep; AbD Serotec), EEA1 (mouse clone 14; BD) and LAMP1 (mouse clone H4A3; BD). Cover slips were mounted on glass slides in Mowiol 4-88.

8.8.3 Image Acquisition

Slides were imaged with a 60×/1.35 NA oil immersion objective on an upright microscope (BX61 Olympus) with filter sets for DAPI, GFP/Alexa 488, 555, 568, and 647 (Chroma Technology Corp.), a camera (PrimΣ; Photometrics), and MetaMorph 7.5 imaging software (Molecular Dynamics Inc.). Illumination was provided by wLS illumination unit (QImaging). Image stacks of 3–5 planes with 0.3 µm spacing through the cell volume were taken. Image stacks were maximum intensity projected and then merged to create 24-bit RGB TIFF files.

8.8.4 Image Analysis

To determine ER retrieval efficiency, the Golgi signal (integrated pixel intensity) for the KDEL receptor was measured in the region defined by the Golgi marker antibody in the presence and absence of KDEL ligand. Golgi signal was normalised to the ER signal, to account for different expression levels. Specifically, images were opened in ImageJ and maximum intensity z-stack projected. A border was drawn around cell in the image with the KDELR-GFP and raw intensity as well as area were measured (*Cell*, *Cell Area*). Another border was drawn around the Golgi in the Golgi marker image. This border was placed into image with KDELR-GFP. Again intensity and area were measured (*Golgi*, *Golgi Area*). The same border was placed in an empty part of the image, intensity and area were measured again (*Golgi Background*). The intensity per area (*Golgi Background* ÷ *Golgi Area*) of the Golgi background was used to calculate the total background intensity of the cell. Then the *Golgi Fraction* = (*Golgi* − *Golgi background*) ÷ (*Cell* − *Cell background*).

To examine the effect of Syb37 on KDELR localisation, line intensity measurements of the different channels were used.

Bibliography

Abergel C. Molecular replacement: tricks and treats. *Acta Crystallogr D Biol Crystallogr*. 2013 Nov; 69(Pt 11):2167–73. doi:10.1107/S0907444913015291.

Abrams EW, Cheng YL, Andrew DJ. Drosophila KDEL receptor function in the embryonic salivary gland and epidermis. *PLoS One*. 2013; 8(10):e77618. doi:10.1371/journal.pone.0077618.

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010 Feb; 66(Pt 2):213–21. doi:10.1107/S0907444909052925.

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*. 2012 Apr; 68(Pt 4):352–67. doi:10.1107/S0907444912001308.

Agback P, Agback T. Direct evidence of a low barrier hydrogen bond in the catalytic triad of a Serine protease. *Sci Rep*. 2018 Jul; 8(1):10078. doi:10.1038/s41598-018-28441-7.

Alanen HI, Raykhel IB, Luukas MJ, Salo KEH, Ruddock LW. Beyond KDEL: the role of positions 5 and 6 in determining ER localization. *J Mol Biol*. 2011 Jun; 409(3):291–7. doi:10.1016/j.jmb.2011.03.070.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 5th ed. Garland Science; 2007.

Aoe T, Cukierman E, Lee A, Cassel D, Peters PJ, Hsu VW. The KDEL receptor, ERD2, regulates intracellular traffic by recruiting a GTPase-activating protein for ARF1. *EMBO J*. 1997 Dec; 16(24):7305–16. doi:10.1093/emboj/16.24.7305.

Aoe T, Lee AJ, van Donselaar E, Peters PJ, Hsu VW. Modulation of intracellular transport by transported proteins: insight from regulation of COPI-mediated transport. *Proc Natl Acad Sci U S A*. 1998 Feb; 95(4):1624–9.

Arakel EC, Richter KP, Clancy A, Schwappach B. δ-COP contains a helix C-terminal to its longin domain key to COPI dynamics and function. *Proc Natl Acad Sci U S A*. 2016 06; 113(25):6916–21. doi:10.1073/pnas.1603544113.

Araki K, Nagata K. Protein folding and quality control in the ER. *Cold Spring Harb Perspect Biol*. 2011 Nov; 3(11):a007526. doi:10.1101/cshperspect.a007526.

Banfield DK. Mechanisms of protein retention in the Golgi. *Cold Spring Harb Perspect Biol*. 2011 Aug; 3(8):a005264. doi:10.1101/cshperspect.a005264.

Bard F, Mazelin L, Péchoux-Longin C, Malhotra V, Jurdic P. Src regulates Golgi structure and KDEL receptor-dependent retrograde transport to the endoplasmic reticulum. *J Biol Chem*. 2003 Nov; 278(47):46601–6. doi:10.1074/jbc.M302221200.

Barlowe C, Orci L, Yeung T, Hosobuchi M, Hamamoto S, Salama N, Rexach MF, Ravazzola M, Amherdt M, Schekman R. COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell*. 1994 Jun; 77(6):895–907. doi:10.1016/0092-8674(94)90138-4.

Barlowe C. Signals for COPII-dependent export from the ER: what's the ticket out? *Trends Cell Biol*. 2003 Jun; 13(6):295–300.

Barlowe C, Helenius A. Cargo Capture and Bulk Flow in the Early Secretory Pathway. *Annu Rev Cell Dev Biol*. 2016 Oct; 32:197–222. doi:10.1146/annurev-cellbio-111315-125016.

Becker B, Blum A, Gießelmann E, Dausend J, Rammo D, Müller NC, Tschacksch E, Steimer M, Spindler J, Becherer U, Rettig J, Breinig F, Schmitt MJ. H/KDEL receptors mediate host cell intoxication by a viral A/B toxin in yeast. *Sci Rep*. 2016; 6:31105. doi:10.1038/srep31105.

Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, Lee T, Pope SH, Riordan GS, Whitlow M. Single-chain antigen-binding proteins. *Science*. 1988 Oct; 242(4877):423–6.

Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vonrhein C, Womack TO. BUSTER. Cambridge, United Kingdom: Global Phasing; 2017.

Bykov YS, Schaffer M, Dodonova SO, Albert S, Plitzko JM, Baumeister W, Engel BD, Briggs JA. The structure of the COPI coat determined within the cell. *Elife*. 2017 11; 6. doi:10.7554/eLife.32493.

Cabrera M, Muñiz M, Hidalgo J, Vega L, Martín ME, Velasco A. The retrieval function of the KDEL receptor requires PKA phosphorylation of its C-terminus. *Mol Biol Cell*. 2003 Oct; 14(10):4114–25. doi:10.1091/mbc.E03-04-0194.

Cancino J, Capalbo A, Di Campli A, Giannotta M, Rizzo R, Jung JE, Di Martino R, Persico M, Heinklein P, Sallese M, Luini A. Control systems of membrane transport at the interface between the endoplasmic reticulum and the Golgi. *Dev Cell*. 2014 Aug; 30(3):280–94. doi:10.1016/j.devcel.2014.06.018.

Cancino J, Jung JE, Luini A. Regulation of Golgi signaling and trafficking by the KDEL receptor. *Histochem Cell Biol*. 2013 Oct; 140(4):395–405. doi:10.1007/s00418-013-1130-9.

de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*. 2006 Jul; 34(Web Server issue):W362–5. doi:10.1093/nar/gkl124.

Chaptal V, Ujwal R, Nie Y, Watanabe A, Kwon S, Abramson J. Fluorescence Detection of Heavy Atom Labeling (FD-HAL): a rapid method for identifying covalently modified cysteine residues by phasing atoms. *J Struct Biol*. 2010 Jul; 171(1):82–7. doi:10.1016/j.jsb.2010.02.005.

Chen L, Liu YC, Tan H, Zhang Y, Xu J, Liu WL, Li ZY, Li WP. Santacruzamate A Ameliorates AD-Like Pathology by Enhancing ER Stress Tolerance Through Regulating the Functions of KDELR and Mia40-ALR in vivo and in vitro. *Front Cell Neurosci*. 2019; 13:61. doi:10.3389/fncel.2019.00061.

Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, Guo WJ, Kim JG, Underwood W, Chaudhuri B, et al. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*. 2010; 468(7323):527.

Cherqui S, Kalatzis V, Trugnan G, Antignac C. The targeting of cystinosin to the lysosomal membrane requires a tyrosine-based signal and a novel sorting motif. *J Biol Chem*. 2001 Apr; 276(16):13314–21. doi:10.1074/jbc.M010562200.

Chitwood PJ, Hegde RS. The Role of EMC during Membrane Protein Biogenesis. *Trends Cell Biol*. 2019 May; 29(5):371–384. doi:10.1016/j.tcb.2019.01.007.

Cleland WW, Kreevoy MM. Low-barrier hydrogen bonds and enzymic catalysis. *Science*. 1994 Jun; 264(5167):1887–90. doi:10.1126/science.8009219.

Clerc S, Hirsch C, Oggier DM, Deprez P, Jakob C, Sommer T, Aebi M. Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum. *J Cell Biol*. 2009 Jan; 184(1):159–72. doi:10.1083/jcb.200809198.

Cosson P, Letourneur F. Coatomer interaction with di-lysine endoplasmic reticulum retention motifs. *Science*. 1994 Mar; 263(5153):1629–31. doi:10.1126/science.8128252.

Cosson P, Perrin J, Bonifacino JS. Anchors aweigh: protein localization and transport mediated by transmembrane domains. *Trends Cell Biol*. 2013 Oct; 23(10):511–7. doi:10.1016/j.tcb.2013.05.005.

Dell'Angelica EC, Bonifacino JS. Coatopathies: Genetic Disorders of Protein Coats. *Annu Rev Cell Dev Biol*. 2019 Aug; doi:10.1146/annurev-cellbio-100818-125234.

Desmyter A, Spinelli S, Roussel A, Cambillau C. Camelid nanobodies: killing two birds with one stone. *Curr Opin Struct Biol*. 2015 Jun; 32:1–8. doi:10.1016/j.sbi.2015.01.001.

Dodonova SO, Diestelkoetter-Bachert P, von Appen A, Hagen WJH, Beck R, Beck M, Wieland F, Briggs JAG. A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science*. 2015 Jul; 349(6244):195–8. doi:10.1126/science.aab1121.

Dodonova SO, Aderhold P, Kopp J, Ganeva I, Röhling S, Hagen WJ, Sinning I, Wieland F, Briggs JA. 9Å structure of the COPI coat reveals that the Arf1 GTPase occupies two contrasting molecular environments. *Elife*. 2017 Jun; 6. doi:10.7554/eLife.26691.

Donaton MCV, Holsbeeks I, Lagatie O, Van Zeebroeck G, Crauwels M, Winderickx J, Thevelein JM. The Gap1 general amino acid permease acts as an amino acid sensor for activation of protein kinase A targets in the yeast Saccharomyces cerevisiae. *Mol Microbiol*. 2003 Nov; 50(3):911–29.

Dumoulin M, Conrath K, Van Meirhaeghe A, Meersman F, Heremans K, Frenken LGJ, Muyldermans S, Wyns L, Matagne A. Single-domain antibody fragments with high conformational stability. *Protein Sci*. 2002 Mar; 11(3):500–15. doi:10.1110/ps.34602.

Ellgaard L, Frickel EM. Calnexin, calreticulin, and ERp57: teammates in glycoprotein folding. *Cell Biochem Biophys*. 2003; 39(3):223–47. doi:10.1385/CBB:39:3:223.

Ellgaard L, Helenius A. Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol*. 2003 Mar; 4(3):181–91. doi:10.1038/nrm1052.

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010 Apr; 66(Pt 4):486–501. doi:10.1107/S0907444910007493.

Ereño-Orbea J, Sicard T, Cui H, Carson J, Hermans P, Julien JP. Structural Basis of Enhanced Crystallizability Induced by a Molecular Chaperone for Antibody Antigen-Binding Fragments. *J Mol Biol*. 2018 Feb; 430(3):322–336. doi:10.1016/j.jmb.2017.12.010.

Evans P, McCoy A. An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr*. 2008 Jan; 64(Pt 1):1–10. doi:10.1107/S0907444907051554.

Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr*. 2011 Apr; 67(Pt 4):282–92. doi:10.1107/S090744491003982X.

Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*. 2013 Jul; 69(Pt 7):1204–14. doi:10.1107/S0907444913000061.

Feng L, Frommer WB. Structure and function of SemiSWEET and SWEET sugar transporters. *Trends Biochem Sci*. 2015 Aug; 40(8):480–6. doi:10.1016/j.tibs.2015.05.005.

Feng L, Frommer WB. Evolution of Transporters: The Relationship of SWEETs, PQ-loop, and PnuC Transporters. *Trends Biochem Sci*. 2016 Feb; 41(2):118–119. doi:10.1016/j.tibs.2015.11.014.

Fridy PC, Li Y, Keegan S, Thompson MK, Nudelman I, Scheid JF, Oeffinger M, Nussenzweig MC, Fenyö D, Chait BT, Rout MP. A robust pipeline for rapid production of versatile nanobody repertoires. *Nat Methods*. 2014 Dec; 11(12):1253–60. doi:10.1038/nmeth.3170.

Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A*. 1999 Aug; 96(17):9459–64. doi:10.1073/pnas.96.17.9459.

Gallivan JP, Dougherty DA. A Computational Study of Cation-π Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering. *Journal of the American Chemical Society*. 2000 Feb; 122(5):870–874. , doi:10.1021/ja991755c.

Geertsma ER. FX cloning: a simple and robust high-throughput cloning method for protein expression. *Methods Mol Biol*. 2014; 1116:153–64. doi:10.1007/978-1-62703-764-8_11.

Giannotta M, Fragassi G, Tamburro A, Vanessa C, Luini A, Sallese M. Prohibitin: A Novel Molecular Player in KDEL Receptor Signalling. *Biomed Res Int*. 2015; 2015:319454. doi:10.1155/2015/319454.

Giannotta M, Ruggiero C, Grossi M, Cancino J, Capitani M, Pulvirenti T, Consoli GML, Geraci C, Fanelli F, Luini A, Sallese M. The KDEL receptor couples to Gαq/11 to activate Src kinases and regulate transport through the Golgi. *EMBO J*. 2012 Jun; 31(13):2869–81. doi:10.1038/emboj.2012.134.

Gomez-Navarro N, Miller E. Protein sorting at the ER-Golgi interface. *J Cell Biol*. 2016 Dec; 215(6):769–778. doi:10.1083/jcb.201610031.

Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. 2006; 1(6):2876–90. doi:10.1038/nprot.2006.202.

Griffiths G, Ericsson M, Krijnse-Locker J, Nilsson T, Goud B, Söling HD, Tang BL, Wong SH, Hong W. Localization of the Lys, Asp, Glu, Leu tetrapeptide receptor to the Golgi complex and the intermediate compartment in mammalian cells. *J Cell Biol*. 1994 Dec; 127(6 Pt 1):1557–74.

Guna A, Hegde RS. Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Curr Biol*. 2018 04; 28(8):R498–R511. doi:10.1016/j.cub.2018.02.004.

Hamada H, Suzuki M, Yuasa S, Mimura N, Shinozuka N, Takada Y, Suzuki M, Nishino T, Nakaya H, Koseki H, Aoe T. Dilated cardiomyopathy caused by aberrant endoplasmic reticulum quality control in mutant KDEL receptor transgenic mice. *Mol Cell Biol*. 2004 Sep; 24(18):8007–17. doi:10.1128/MCB.24.18.8007-8017.2004.

Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, Bendahman N, Hamers R. Naturally occurring antibodies devoid of light chains. *Nature*. 1993 Jun; 363(6428):446–8. doi:10.1038/363446a0.

Hammond C, Helenius A. Quality control in the secretory pathway: retention of a misfolded viral membrane glycoprotein involves cycling between the ER, intermediate compartment, and Golgi apparatus. *J Cell Biol*. 1994 Jul; 126(1):41–52.

Harris TK, Mildvan AS. High-precision measurement of hydrogen bond lengths in proteins by nuclear magnetic resonance methods. *Proteins*. 1999 May; 35(3):275–82.

Hegde RS, Keenan RJ. Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat Rev Mol Cell Biol*. 2011 Nov; 12(12):787–98. doi:10.1038/nrm3226.

Heger A, Holm L. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins*. 2000 Nov; 41(2):224–37. doi:10.1002/1097-0134(20001101)41:2<224::aid-prot70>3.0.co;2-z.

von Heijne G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol*. 1992 May; 225(2):487–94. doi:10.1016/0022-2836(92)90934-c.

Helms JB, Helms-Brons D, Brügger B, Gkantiragas I, Eberle H, Nickel W, Nürnberg B, Gerdes HH, Wieland FT. A putative heterotrimeric G protein inhibits the fusion of COPI-coated vesicles. Segregation of heterotrimeric G proteins from COPI-coated vesicles. *J Biol Chem*. 1998 Jun; 273(24):15203–8.

Hessler JL, Kreitman RJ. An early step in Pseudomonas exotoxin action is removal of the terminal lysine residue, which allows binding to the KDEL receptor. *Biochemistry*. 1997 Nov; 36(47):14577–82. doi:10.1021/bi971447w.

Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A. Moving beyond P values: data analysis with estimation graphics. *Nat Methods*. 2019 07; 16(7):565–566. doi:10.1038/s41592-019-0470-3.

Holzwarth G, Doty P. The Ultraviolet Circular Dichroism of Polypeptides. *J Am Chem Soc*. 1965 Jan; 87:218–28. doi:10.1021/ja01080a015.

Hsu VW, Yuan LC, Nuchtern JG, Lippincott-Schwartz J, Hammerling GJ, Klausner RD. A recycling pathway between the endoplasmic reticulum and the Golgi apparatus for retention of unassembled MHC class I molecules. *Nature*. 1991 Aug; 352(6334):441–4. doi:10.1038/352441a0.

Hyde R, Cwiklinski EL, MacAulay K, Taylor PM, Hundal HS. Distinct sensor pathways in the hierarchical control of SNAT2, a putative amino acid transceptor, by amino acid availability. *J Biol Chem*. 2007 Jul; 282(27):19788–98. doi:10.1074/jbc.M611520200.

Jackson LP, Lewis M, Kent HM, Edeling MA, Evans PR, Duden R, Owen DJ. Molecular basis for recognition of dilysine trafficking motifs by COPI. *Dev Cell*. 2012 Dec; 23(6):1255–62. doi:10.1016/j.devcel.2012.10.017.

Jackson ME, Simpson JC, Girod A, Pepperkok R, Roberts LM, Lord JM. The KDEL retrieval system is exploited by Pseudomonas exotoxin A, but not by Shiga-like toxin-1, during retrograde transport from the Golgi complex to the endoplasmic reticulum. *J Cell Sci*. 1999 Feb; 112 (Pt 4):467–75.

Jaehme M, Guskov A, Slotboom DJ. Crystal structure of the vitamin B3 transporter PnuC, a full-length SWEET homolog. *Nat Struct Mol Biol*. 2014 Nov; 21(11):1013–5. doi:10.1038/nsmb.2909.

Jaehme M, Guskov A, Slotboom DJ. The twisted relation between Pnu and SWEET transporters. *Trends Biochem Sci*. 2015 Apr; 40(4):183–8. doi:10.1016/j.tibs.2015.02.002.

Jaehme M, Guskov A, Slotboom DJ. Pnu Transporters: Ain't They SWEET? *Trends Biochem Sci*. 2016 Feb; 41(2):117–118. doi:10.1016/j.tibs.2015.11.013.

Jazayeri A, Rappas M, Brown AJH, Kean J, Errey JC, Robertson NJ, Fiez-Vandal C, Andrews SP, Congreve M, Bortolato A, Mason JS, Baig AH, Teobald I, Doré AS, Weir M, Cooke RM, Marshall FH. Crystal structure of the GLP-1 receptor bound to a peptide agonist. *Nature*. 2017 06; 546(7657):254–258. doi:10.1038/nature22800.

Jeffrey GA. An introduction to hydrogen bonding, vol. 32. Oxford university press New York; 1997.

Jézégou A, Llinares E, Anne C, Kieffer-Jaquinod S, O'Regan S, Aupetit J, Chabli A, Sagné C, Debacker C, Chadefaux-Vekemans B, Journet A, André B, Gasnier B. Heptahelical protein PQLC2 is a lysosomal cationic amino acid exporter underlying the action of cysteamine in cystinosis therapy. *Proc Natl Acad Sci U S A*. 2012 Dec; 109(50):E3434–43. doi:10.1073/pnas.1211198109.

Jia B, Zhu XF, Pu ZJ, Duan YX, Hao LJ, Zhang J, Chen LQ, Jeon CO, Xuan YH. Integrative View of the Diversity and Evolution of SWEET and SemiSWEET Sugar Transporters. *Front Plant Sci*. 2017; 8:2178. doi:10.3389/fpls.2017.02178.

Jin H, Komita M, Aoe T. The Role of BiP Retrieval by the KDEL Receptor in the Early Secretory Pathway and its Effect on Protein Quality Control and Neurodegeneration. *Front Mol Neurosci*. 2017; 10:222. doi:10.3389/fnmol.2017.00222.

Jin H, Mimura N, Kashio M, Koseki H, Aoe T. Late-onset of spinal neurodegeneration in knock-in mice expressing a mutant BiP. *PLoS One*. 2014; 9(11):e112837. doi:10.1371/journal.pone.0112837.

Johnson N, Powis K, High S. Post-translational translocation into the endoplasmic reticulum. *Biochim Biophys Acta*. 2013 Nov; 1833(11):2403–9. doi:10.1016/j.bbamcr.2012.12.008.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*. 1991 Mar; 47 (Pt 2):110–9.

Jungnickel KEJ, Parker JL, Newstead S. Structural basis for amino acid transport by the CAT family of SLC7 transporters. *Nat Commun*. 2018 02; 9(1):550. doi:10.1038/s41467-018-03066-6.

Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-Gordon T, Geng W, Krasny R, et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci*. 2018 01; 27(1):112–128. doi:10.1002/pro.3280.

Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983 Dec; 22(12):2577–637. doi:10.1002/bip.360221211.

Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010 Feb; 66(Pt 2):125–32. doi:10.1107/S0907444909047337.

Kalatzis V, Cherqui S, Antignac C, Gasnier B. Cystinosin, the protein defective in cystinosis, is a H(+)-driven lysosomal cystine transporter. *EMBO J*. 2001 Nov; 20(21):5940–9. doi:10.1093/emboj/20.21.5940.

Kamimura D, Atsumi T, Stofkova A, Nishikawa N, Ohki T, Suzuki H, Katsunuma K, Jiang JJ, Bando H, Meng J, Sabharwal L, Ogura H, Hirano T, Arima Y, Murakami M. Naïve T Cell Homeostasis Regulated by Stress Responses and TCR Signaling. *Front Immunol*. 2015; 6:638. doi:10.3389/fimmu.2015.00638.

Kanapin A, Batalov S, Davis MJ, Gough J, Grimmond S, Kawaji H, Magrane M, Matsuda H, Schönbach C, Teasdale RD, Group RG, Members G, Yuan3 Z. Mouse Proteome Analysis. *Genome Research*. 2003; 13(6b):1335–1344. , doi:10.1101/gr.978703.

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000 Jan; 28(1):27–30. doi:10.1093/nar/28.1.27.

Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure*. 2006 Apr; 14(4):673–81. doi:10.1016/j.str.2006.01.013.

Keyes MH, Gray DN, Kreh KE, Sanders CR. Solubilizing Detergents for Membrane Proteins. In: Iwata S, editor. *Methods and Results in Crystallization of Membrane Proteins*, vol. 4 of IUL Biotechnology Series International University Line; 2003.p. 17–38.

Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, Them NCC, Berg T, Gisslinger B, Pietra D, Chen D, Vladimer GI, Bagienski K, Milanesi C, Casetti IC, Sant'Antonio E, Ferretti V, Elena C, Schischlik F, Cleary C, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med*. 2013 Dec; 369(25):2379–90. doi:10.1056/NEJMoa1311347.

Koide S. Engineering of recombinant crystallization chaperones. *Curr Opin Struct Biol*. 2009 Aug; 19(4):449–57. doi:10.1016/j.sbi.2009.04.008.

Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007 Sep; 372(3):774–97. doi:10.1016/j.jmb.2007.05.022.

Latorraca NR, Fastman NM, Venkatakrishnan AJ, Frommer WB, Dror RO, Feng L. Mechanism of Substrate Translocation in an Alternating Access Transporter. *Cell*. 2017 03; 169(1):96–107.e12. doi:10.1016/j.cell.2017.03.010.

Lee S, Mao A, Bhattacharya S, Robertson N, Grisshammer R, Tate CG, Vaidehi N. How Do Short Chain Nonionic Detergents Destabilize G-Protein-

Coupled Receptors? *J Am Chem Soc*. 2016 11; 138(47):15425–15433. doi:10.1021/jacs.6b08742.

Lee Y, Nishizawa T, Yamashita K, Ishitani R, Nureki O. Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. *Nat Commun*. 2015 Jan; 6:6112. doi:10.1038/ncomms7112.

Lesser GJ, Rose GD. Hydrophobicity of amino acid subgroups in proteins. *Proteins*. 1990; 8(1):6–13. doi:10.1002/prot.340080104.

Letourneur F, Gaynor EC, Hennecke S, Démollière C, Duden R, Emr SD, Riezman H, Cosson P. Coatomer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. *Cell*. 1994 Dec; 79(7):1199–207. doi:10.1016/0092-8674(94)90011-6.

Lewis MJ, Pelham HR. A human homologue of the yeast HDEL receptor. *Nature*. 1990 Nov; 348(6297):162–3. doi:10.1038/348162a0.

Lewis MJ, Pelham HR. Ligand-induced redistribution of a human KDEL receptor from the Golgi complex to the endoplasmic reticulum. *Cell*. 1992 Jan; 68(2):353–64.

Lewis MJ, Pelham HR. Sequence of a second human KDEL receptor. *J Mol Biol*. 1992 Aug; 226(4):913–6.

Leyte A, Barr FA, Kehlenbach RH, Huttner WB. Multiple trimeric G-proteins on the trans-Golgi network exert stimulatory and inhibitory effects on secretory vesicle formation. *EMBO J*. 1992 Dec; 11(13):4795–804.

Li MY, Grandadam M, Kwok K, Lagache T, Siu YL, Zhang JS, Sayteng K, Kudelko M, Qin CF, Olivo-Marin JC, Bruzzone R, Wang PG. KDEL Receptors Assist Dengue Virus Exit from the Endoplasmic Reticulum. *Cell Rep*. 2015 Mar; doi:10.1016/j.celrep.2015.02.021.

Liao SM, Du QS, Meng JZ, Pang ZW, Huang RB. The multiple roles of histidine in protein interactions. *Chem Cent J*. 2013 Mar; 7(1):44. doi:10.1186/1752-153X-7-44.

Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, Adams PD. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol*. 2017 02; 73(Pt 2):148–157. doi:10.1107/S2059798316018210.

Lin JH, Walter P, Yen TSB. Endoplasmic reticulum stress in disease pathogenesis. *Annu Rev Pathol*. 2008; 3:399–425. doi:10.1146/annurev.pathmechdis.3.121806.151434.

Liu B, Du H, Rutkowski R, Gartner A, Wang X. LAAT-1 is the lysosomal lysine/arginine transporter that maintains amino acid homeostasis. *Science*. 2012 Jul; 337(6092):351–4. doi:10.1126/science.1220281.

Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J. Molecular Structure of the Human CFTR Ion Channel. *Cell*. 2017 03; 169(1):85–95.e8. doi:10.1016/j.cell.2017.02.024.

Liu L, Doray B, Kornfeld S. Recycling of Golgi glycosyltransferases requires direct binding to coatomer. *Proc Natl Acad Sci U S A*. 2018 09; 115(36):8984–8989. doi:10.1073/pnas.1810291115.

Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*. 2003 Feb; 50(3):437–50. doi:10.1002/prot.10286.

Ma W, Goldberg J. Rules for the recognition of dilysine retrieval motifs by coatomer. *EMBO J*. 2013 Apr; 32(7):926–37. doi:10.1038/emboj.2013.41.

Majoul I, Straub M, Hell SW, Duden R, Söling HD. KDEL-cargo regulates interactions between proteins involved in COPI vesicle traffic: measurements in living cells using FRET. *Dev Cell*. 2001 Jul; 1(1):139–53.

Mancias JD, Goldberg J. The transport signal on Sec22 for packaging into COPII-coated vesicles is a conformational epitope. *Mol Cell*. 2007 May; 26(3):403–14. doi:10.1016/j.molcel.2007.03.017.

Mancias JD, Goldberg J. Structural basis of cargo membrane protein discrimination by the human COPII coat machinery. *EMBO J*. 2008 Nov; 27(21):2918–28. doi:10.1038/emboj.2008.208.

Matlack KE, Misselwitz B, Plath K, Rapoport TA. BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane. *Cell*. 1999 May; 97(5):553–64. doi:10.1016/s0092-8674(00)80767-9.

McCoy AJ. Liking likelihood. *Acta Crystallogr D Biol Crystallogr*. 2004 Dec; 60(Pt 12 Pt 1):2169–83. doi:10.1107/S0907444904016038.

McCoy AJ. Acknowledging Errors: Advanced Molecular Replacement with Phaser. *Methods Mol Biol*. 2017; 1607:421–453. doi:10.1007/978-1-4939-7000-1_18.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007 08; 40(Pt 4):658–674. doi:10.1107/S0021889807021206.

van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol*. 2008 Feb; 9(2):112–24. doi:10.1038/nrm2330.

Mei M, Zhai C, Li X, Zhou Y, Peng W, Ma L, Wang Q, Iverson BL, Zhang G, Yi L. Characterization of aromatic residue-controlled protein retention in the endoplasmic reticulum of Saccharomyces cerevisiae. *J Biol Chem*. 2017 Dec; 292(50):20707–20719. doi:10.1074/jbc.M117.812107.

Mihaesco C, Seligmann M. Papain digestion fragments of human IgM globulins. *J Exp Med*. 1968 Mar; 127(3):431–53.

Mimura N, Hamada H, Kashio M, Jin H, Toyama Y, Kimura K, Iida M, Goto S, Saisho H, Toshimori K, Koseki H, Aoe T. Aberrant quality control in the endoplasmic reticulum impairs the biosynthesis of pulmonary surfactant in mice expressing mutant BiP. *Cell Death Differ*. 2007 Aug; 14(8):1475–85. doi:10.1038/sj.cdd.4402151.

Mimura N, Yuasa S, Soma M, Jin H, Kimura K, Goto S, Koseki H, Aoe T. Altered quality control in the endoplasmic reticulum causes cortical dysplasia in knock-in mice expressing a mutant BiP. *Mol Cell Biol*. 2008 Jan; 28(1):293–301. doi:10.1128/MCB.00473-07.

Munro S, Pelham HR. A C-terminal signal prevents secretion of luminal ER proteins. *Cell*. 1987 Mar; 48(5):899–907. doi:10.1016/0092-8674(87)90086-9.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2011 Apr; 67(Pt 4):355–67. doi:10.1107/S0907444911001314.

Muyldermans S. Nanobodies: natural single-domain antibodies. *Annu Rev Biochem*. 2013; 82:775–97. doi:10.1146/annurev-biochem-063011-092449.

Netherton C, Rouiller I, Wileman T. The subcellular distribution of multigene family 110 proteins of African swine fever virus is determined by differences in C-terminal KDEL endoplasmic reticulum retention motifs. *J Virol*. 2004 Apr; 78(7):3710–21. doi:10.1128/JVI.78.7.3710-3721.2004.

Newstead S, Kim H, von Heijne G, Iwata S, Drew D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 2007 Aug; 104(35):13936–41. doi:10.1073/pnas.0704546104.

Nishimura N, Balch WE. A di-acidic signal required for selective export from the endoplasmic reticulum. *Science*. 1997 Jul; 277(5325):556–8.

Oakes SA, Papa FR. The role of endoplasmic reticulum stress in human pathology. *Annu Rev Pathol*. 2015; 10:173–94. doi:10.1146/annurev-pathol-012513-104649.

Oeffner RD, Afonine PV, Millán C, Sammito M, Usón I, Read RJ, McCoy AJ. On the application of the expected log-likelihood gain to decision making in molecular replacement. *Acta Crystallogr D Struct Biol*. 2018 04; 74(Pt 4):245–255. doi:10.1107/S2059798318004357.

Parker JL, Newstead S. Phasing statistics for alpha helical membrane protein structures. *Protein Sci*. 2013 Nov; 22(11):1664–8. doi:10.1002/pro.2341.

Parker JL, Newstead S. Method to increase the yield of eukaryotic membrane protein expression in Saccharomyces cerevisiae for structural and functional studies. *Protein Sci*. 2014 Sep; 23(9):1309–14. doi:10.1002/pro.2507.

Parker JL, Newstead S. Membrane Protein Crystallisation: Current Trends and Future Perspectives. In: Morales I, editor. *The Next Generation in Membrane Protein Structure Determination*, vol. 922 of Advances in Experimental Medicine and Biology Springer International Publishing; 2016.p. 61–72. doi:10.1007/978-3-319-35072-1.

Parker JL, Newstead S. Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature*. 2017 11; 551(7681):521–524. doi:10.1038/nature24464.

Paroutis P, Touret N, Grinstein S. The pH of the secretory pathway: measurement, determinants, and regulation. *Physiology (Bethesda)*. 2004 Aug; 19:207–15. doi:10.1152/physiol.00005.2004.

Pelham HR. Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment. *EMBO J*. 1988 Apr; 7(4):913–8.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem*. 2004 Oct; 25(13):1605–12. doi:10.1002/jcc.20084.

Phillips MJ, Voeltz GK. Structure and function of ER membrane contact sites with other organelles. *Nat Rev Mol Cell Biol*. 2016 Feb; 17(2):69–82. doi:10.1038/nrm.2015.8.

Pimplikar SW, Simons K. Regulation of apical transport in epithelial cells by a Gs class of heterotrimeric G protein. *Nature*. 1993 Apr; 362(6419):456–8. doi:10.1038/362456a0.

Piovesan D, Minervini G, Tosatto SCE. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res*. 2016 07; 44(W1):W367–74. doi:10.1093/nar/gkw315.

Privé GG. Detergents for the stabilization and crystallization of membrane proteins. *Methods*. 2007 Apr; 41(4):388–97. doi:10.1016/j.ymeth.2007.01.007.

Pulvirenti T, Giannotta M, Capestrano M, Capitani M, Pisanu A, Polishchuk RS, Pietro ES, Beznoussenko GV, Mironov AA, Turacchio G, Hsu VW, Sallese M, Luini A. A traffic-activated Golgi-based signalling circuit coordinates the secretory pathway. *Nat Cell Biol*. 2008 08; 10(8):912–922. .

Qi HW, Kulik HJ. Evaluating Unexpectedly Short Non-covalent Distances in X-ray Crystal Structures of Proteins with Electronic Structure Analysis. *J Chem Inf Model*. 2019 Mar; doi:10.1021/acs.jcim.9b00144.

Rapoport TA, Li L, Park E. Structural and Mechanistic Insights into Protein Translocation. *Annu Rev Cell Dev Biol*. 2017 10; 33:369–390. doi:10.1146/annurev-cellbio-100616-060439.

Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK. Crystal structure of the β2 adrenergic receptor-Gs protein complex. *Nature*. 2011 Jul; 477(7366):549–55. doi:10.1038/nature10361.

Raykhel I, Alanen H, Salo K, Jurvansuu J, Nguyen VD, Latva-Ranta M, Ruddock L. A molecular specificity code for the three mammalian KDEL receptors. *J Cell Biol*. 2007 Dec; 179(6):1193–204. doi:10.1083/jcb.200705180.

Ruggiano A, Foresti O, Carvalho P. Quality control: ER-associated degradation: protein quality control and beyond. *J Cell Biol*. 2014 Mar; 204(6):869–79. doi:10.1083/jcb.201312042.

Sato K, Nakano A. Mechanisms of COPII vesicle formation and protein sorting. *FEBS Lett*. 2007 May; 581(11):2076–82. doi:10.1016/j.febslet.2007.01.091.

Sato K, Sato M, Nakano A. Rer1p, a retrieval receptor for ER membrane proteins, recognizes transmembrane domains in multiple modes. *Mol Biol Cell*. 2003 Sep; 14(9):3605–16. doi:10.1091/mbc.e02-12-0777.

Saudek V. Cystinosin, MPDU1, SWEETs and KDELR belong to a well-defined protein family with putative function of cargo receptors involved in vesicle trafficking. *PLoS One*. 2012; 7(2):e30876. doi:10.1371/journal.pone.0030876.

Saudek V. Potential dual function of PQ-loop proteins such as cystinosin. *J Biol Chem*. 2017 09; 292(36):15133. doi:10.1074/jbc.L117.795278.

Scheel AA, Pelham HR. Purification and characterization of the human KDEL receptor. *Biochemistry*. 1996 Aug; 35(31):10203–9. doi:10.1021/bi960807x.

Scheel AA, Pelham HR. Identification of amino acids in the binding pocket of the human KDEL receptor. *J Biol Chem*. 1998 Jan; 273(4):2467–72.

Schomaker V, Trueblood K. On the rigid-body motion of molecules in crystals. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*. 1968; 24(1):63–76.

Seetharam S, Chaudhary VK, FitzGerald D, Pastan I. Increased cytotoxic activity of Pseudomonas exotoxin and two chimeric toxins ending in KDEL. *J Biol Chem*. 1991 Sep; 266(26):17376–81.

Semenza JC, Pelham HR. Changing the specificity of the sorting receptor for luminal endoplasmic reticulum proteins. *J Mol Biol*. 1992 Mar; 224(1):1–5.

Semenza JC, Hardwick KG, Dean N, Pelham HR. ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. *Cell*. 1990; 61(7):1349–1357. doi:10.1016/0092-8674(90)90698-E.

Shapovalov MV, Dunbrack RL Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*. 2011 Jun; 19(6):844–58. doi:10.1016/j.str.2011.03.019.

Sharpe HJ, Stevens TJ, Munro S. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell*. 2010 Jul; 142(1):158–69. doi:10.1016/j.cell.2010.05.037.

Shibuya A, Margulis N, Christiano R, Walther TC, Barlowe C. The Erv41-Erv46 complex serves as a retrograde receptor to retrieve escaped ER proteins. *J Cell Biol*. 2015 Jan; 208(2):197–209. doi:10.1083/jcb.201408024.

Siggs OM, Popkin DL, Krebs P, Li X, Tang M, Zhan X, Zeng M, Lin P, Xia Y, Oldstone MBA, Cornall RJ, Beutler B. Mutation of the ER retention receptor KDELR1 leads to cell-intrinsic lymphopenia and a failure to control chronic viral infection. *Proc Natl Acad Sci U S A*. 2015 Oct; 112(42):E5706–14. doi:10.1073/pnas.1515619112.

Singh P, Tang BL, Wong SH, Hong W. Transmembrane topology of the mammalian KDEL receptor. *Mol Cell Biol*. 1993 Oct; 13(10):6435–41.

Solis GP, Bilousov O, Koval A, Lüchtenborg AM, Lin C, Katanaev VL. Golgi-Resident Gαo Promotes Protrusive Membrane Dynamics. *Cell*. 2017 09; 170(6):1258. doi:10.1016/j.cell.2017.08.045.

Sonoda Y, Newstead S, Hu NJ, Alguel Y, Nji E, Beis K, Yashiro S, Lee C, Leung J, Cameron AD, Byrne B, Iwata S, Drew D. Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure*. 2011 Jan; 19(1):17–25. doi:10.1016/j.str.2010.12.001.

Stetsenko A, Guskov A. An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization. *Crystals*. 2017 July; 7(7). doi:10.3390/cryst7070197.

Stickle DF, Presta LG, Dill KA, Rose GD. Hydrogen bonding in globular proteins. *J Mol Biol*. 1992 Aug; 226(4):1143–59. doi:10.1016/0022-2836(92)91058-w.

Stokes AJ, Wakano C, del Carmen KA, Koblan-Huberson M, Turner H. Formation of a physiological complex between TRPV2 and RGA protein promotes cell surface expression of TRPV2. *Journal of cellular biochemistry*. 2005; 94(4):669–683.

Stow JL, de Almeida JB, Narula N, Holtzman EJ, Ercolani L, Ausiello DA. A heterotrimeric G protein, G alpha i-3, on Golgi membranes regulates the secretion of a heparan sulfate proteoglycan in LLC-PK1 epithelial cells. *J Cell Biol*. 1991 Sep; 114(6):1113–24.

Takagi M, Ishii T, Barnes AM, Weis M, Amano N, Tanaka M, Fukuzawa R, Nishimura G, Eyre DR, Marini JC, Hasegawa T. A novel mutation in LEPRE1 that eliminates only the KDEL ER- retrieval sequence causes non-lethal osteogenesis imperfecta. *PLoS One*. 2012; 7(5):e36809. doi:10.1371/journal.pone.0036809.

Tao Y, Cheung LS, Li S, Eom JS, Chen LQ, Xu Y, Perry K, Frommer WB, Feng L. Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature*. 2015 Nov; 527(7577):259–63. doi:10.1038/nature15391.

Tapia D, Jiménez T, Zamora C, Espinoza J, Rizzo R, González-Cárdenas A, Fuentes D, Hernández S, Cavieres VA, Soza A, Guzmán F, Arriagada G, Yuseff MI, Mardones GA, Burgos PV, Luini A, González A, Cancino J. KDEL receptor regulates secretion by lysosome relocation- and autophagy-dependent modulation of lipid-droplet turnover. *Nature Communications*. 2019; 10(1):735. , doi:10.1038/s41467-019-08501-w.

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr*. 2008 Jan; 64(Pt 1):61–9. doi:10.1107/S090744490705024X.

Thompson AA, Liu JJ, Chun E, Wacker D, Wu H, Cherezov V, Stevens RC. GPCR stabilization using the bicelle-like architecture of mixed sterol-detergent micelles. *Methods*. 2011 Dec; 55(4):310–7. doi:10.1016/j.ymeth.2011.10.011.

Thor F, Gautschi M, Geiger R, Helenius A. Bulk flow revisited: transport of a soluble protein in the secretory pathway. *Traffic*. 2009 Dec; 10(12):1819–30. doi:10.1111/j.1600-0854.2009.00989.x.

Tiwarekar V, Fehrholz M, Schneider-Schaulies J. KDELR2 Competes with Measles Virus Envelope Proteins for Cellular Chaperones Reducing Their Chaperone-Mediated Cell Surface Transport. *Viruses*. 2019 Jan; 11(1). doi:10.3390/v11010027.

Townsley FM, Frigerio G, Pelham HR. Retrieval of HDEL proteins is required for growth of yeast cells. *J Cell Biol*. 1994 Oct; 127(1):21–8.

Townsley FM, Wilson DW, Pelham HR. Mutational analysis of the human KDEL receptor: distinct structural requirements for Golgi retention, ligand binding and retrograde transport. *EMBO J*. 1993 Jul; 12(7):2821–9.

Trychta KA, Bäck S, Henderson MJ, Harvey BK. KDEL Receptors Are Differentially Regulated to Maintain the ER Proteome under Calcium Deficiency. *Cell Rep*. 2018 Nov; 25(7):1829–1840.e6. doi:10.1016/j.celrep.2018.10.055.

UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019 Jan; 47(D1):D506–D515. doi:10.1093/nar/gky1049.

Urzhumtsev A, Afonine PV, Adams PD. TLS for dummies. *Comput Crystallogr Newsl*. 2011; 2:42–84.

Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr*. 1999 Jan; 55(Pt 1):191–205. doi:10.1107/S0907444998006684.

Vilardi F, Stephan M, Clancy A, Janshoff A, Schwappach B. WRB and CAML are necessary and sufficient to mediate tail-anchored protein targeting to the ER membrane. *PLoS One*. 2014; 9(1):e85033. doi:10.1371/journal.pone.0085033.

Walden H. Selenium incorporation using recombinant techniques. *Acta Crystallogr D Biol Crystallogr*. 2010 Apr; 66(Pt 4):352–7. doi:10.1107/S0907444909038207.

Wang P, Li B, Zhou L, Fei E, Wang G. The KDEL receptor induces autophagy to promote the clearance of neurodegenerative disease-related proteins. *Neuroscience*. 2011 Sep; 190:43–55. doi:10.1016/j.neuroscience.2011.06.008.

Wang RYL, Wu YJ, Chen HS, Chen CJ. A KDEL Retrieval System for ER-Golgi Transport of Japanese Encephalitis Viral Particles. *Viruses*. 2016 Feb; 8(2). doi:10.3390/v8020044.

Wang X, Matteson J, An Y, Moyer B, Yoo JS, Bannykh S, Wilson IA, Riordan JR, Balch WE. COPII-dependent export of cystic fibrosis transmembrane conductance regulator from the ER uses a di-acidic exit code. *J Cell Biol*. 2004 Oct; 167(1):65–74. doi:10.1083/jcb.200401035.

Watanabe S, Harayama M, Kanemura S, Sitia R, Inaba K. Structural basis of pH-dependent client binding by ERp44, a key regulator of protein secretion at the ER-Golgi interface. *Proc Natl Acad Sci U S A*. 2017 04; 114(16):E3224–E3232. doi:10.1073/pnas.1621426114.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018 07; 46(W1):W296–W303. doi:10.1093/nar/gky427.

Waters MG, Serafini T, Rothman JE. 'Coatomer': a cytosolic protein complex containing subunits of non-clathrin-coated Golgi transport vesicles. *Nature*. 1991 Jan; 349(6306):248–51. doi:10.1038/349248a0.

White JF, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, Gvozdenovic-Jeremic J, Shah P, Shiloach J, Tate CG, Grisshammer R. Structure of the agonist-bound neurotensin receptor. *Nature*. 2012 Oct; 490(7421):508–13. doi:10.1038/nature11558.

Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB 3rd, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*. 2018 01; 27(1):293–315. doi:10.1002/pro.3330.

Wilson CA, Kreychman J, Gerstein M. Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. *J Mol Biol*. 2000 Mar; 297(1):233–49. doi:10.1006/jmbi.2000.3550.

Wilson DW, Lewis MJ, Pelham HR. pH-dependent binding of KDEL to its receptor in vitro. *J Biol Chem*. 1993 Apr; 268(10):7465–8.

Winter G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr*. 2010 February; 43:186–190. doi:10.1107/S0021889809045701.

Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, Fuentes-Montero L, Vollmar M, Michels-Clark T, Young ID, Sauter NK, Evans G. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol*. 2018 02; 74(Pt 2):85–97. doi:10.1107/S2059798317017235.

Wu F, Song G, de Graaf C, Stevens RC. Structure and Function of Peptide-Binding G Protein-Coupled Receptors. *J Mol Biol*. 2017 08; 429(17):2726–2745. doi:10.1016/j.jmb.2017.06.022.

Wu MM, Grabe M, Adams S, Tsien RY, Moore HP, Machen TE. Mechanisms of pH regulation in the regulated secretory pathway. *J Biol Chem*. 2001 Aug; 276(35):33027–35. doi:10.1074/jbc.M103917200.

Wu MM, Llopis J, Adams S, McCaffery JM, Kulomaa MS, Machen TE, Moore HP, Tsien RY. Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem Biol*. 2000 Mar; 7(3):197–209.

Yamaguchi S, Kamikubo H, Kurihara K, Kuroki R, Niimura N, Shimizu N, Yamazaki Y, Kataoka M. Low-barrier hydrogen bond in photoactive yellow protein. *Proc Natl Acad Sci U S A*. 2009 Jan; 106(2):440–4. doi:10.1073/pnas.0811882106.

Yamamoto K, Fujii R, Toyofuku Y, Saito T, Koseki H, Hsu VW, Aoe T. The KDEL receptor mediates a retrieval mechanism that contributes to quality control at the endoplasmic reticulum. *The EMBO Journal*. 2001; 20(12):3082–3091. , doi:10.1093/emboj/20.12.3082.

Yamamoto K, Hamada H, Shinkai H, Kohno Y, Koseki H, Aoe T. The KDEL receptor modulates the endoplasmic reticulum stress response through mitogen-activated protein kinase signaling cascades. *J Biol Chem*. 2003 Sep; 278(36):34525–32. doi:10.1074/jbc.M304188200.

Yamamoto T, Fujimura-Kamada K, Shioji E, Suzuki R, Tanaka K. Cfs1p, a Novel Membrane Protein in the PQ-Loop Family, Is Involved in Phospholipid Flippase Functions in Yeast. *G3 (Bethesda)*. 2017 01; 7(1):179–192. doi:10.1534/g3.116.035238.

Yang S, Wu Y, Xu TH, de Waal PW, He Y, Pu M, Chen Y, DeBruine ZJ, Zhang B, Zaidi SA, Popov P, Guo Y, Han GW, Lu Y, Suino-Powell K, Dong S, Harikumar KG, Miller LJ, Katritch V, Xu HE, et al. Crystal structure of the Frizzled 4 receptor in a ligand-free state. *Nature*. 2018 08; 560(7720):666–670. doi:10.1038/s41586-018-0447-x.

Yee DC, Shlykov MA, Västermark A, Reddy VS, Arora S, Sun EI, Saier MH Jr. The transporter-opsin-G protein-coupled receptor (TOG) superfamily. *FEBS J*. 2013 Nov; 280(22):5780–800. doi:10.1111/febs.12499.

Zanetti G, Prinz S, Daum S, Meister A, Schekman R, Bacia K, Briggs JAG. The structure of the COPII transport-vesicle coat assembled on membranes. *Elife*. 2013 Sep; 2:e00951. doi:10.7554/eLife.00951.

Zhang J, Johnson JL, He J, Napolitano G, Ramadass M, Rocca C, Kiosses WB, Bucci C, Xin Q, Gavathiotis E, Cuervo AM, Cherqui S, Catz SD. Cystinosin, the small GTPase Rab11, and the Rab7 effector RILP regulate intracellular trafficking of the chaperone-mediated autophagy receptor LAMP2A. *J Biol Chem*. 2017 06; 292(25):10328–10346. doi:10.1074/jbc.M116.764076.

Zimmermann I, Egloff P, Hutter CA, Arnold FM, Stohler P, Bocquet N, Hug MN, Huber S, Siegrist M, Hetemann L, Gera J, Gmür S, Spies P, Gygax D, Geertsma ER, Dawson RJ, Seeger MA. Synthetic single domain antibodies for the conformational trapping of membrane proteins. *Elife*. 2018 May; 7. doi:10.7554/eLife.34317.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol*. 2018 07; 430(15):2237–2243. doi:10.1016/j.jmb.2017.12.007.

Acronyms & Abbreviations

A

ARF GAP1 ADP-ribosylation factor GTPase-activating protein 1
At *Arabidopsis thaliana*
ATP adenosine triphosphate
BiP binding immunoglobulin protein
BLI Bio-layer Interferometry
B-PER II Bacterial Protein Extraction Reagent (2X)
BSA Bovine Serum Albumin
CD Circular Dichroism
cDNA complimentary DNA
CDR Complementarity Determining Region
Ce *Caenorhabditis elegans*
CHAPS 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate]
CHS cholesteryl hemisuccinate
CMC Critical Micelle Concentration
CMV cytomegalovirus
COPI coat protein complex I
COPII coat protein complex II
Ct *Chaetomium thermophilum*
CFTR cystic fibrosis transmembrane conductance regulator
CYMAL-5 5-cyclohexyl-1-pentyl-β-D-maltoside
DDM n-dodecyl β-D-maltoside
DHPC 1,2-diheptanoyl-sn-glycero-3-phosphocholine
DM n-decyl-β-D-maltopyranoside
DMEM Dulbecco's Modified Eagle's Medium
DMNG 2,2-dioctylpropane-1,3-bis-β-D-maltopyranoside
DMSO dimethyl sulfoxide
DNA deoxyribonucleic acid
EDTA ethylenediaminetetraacetic acid
EF1α Human elongation factor-1 α
eGFP enhanced GFP
ELISA Enzyme-Linked Immunosorbent Assay
EMC ER membrane protein complex
ER Endoplasmic Reticulum
ERAD ER-associated degradation
ERD ER retention deficient
ERES ER exit sites
ERGIC ER-Golgi intermediate compartment

Fab antigen binding fragment
FD-HAL Fluorescence Detection of Heavy Atom Labelling
FP fluorescence polarisation
FRET Förster resonance energy transfer
FSEC Fluorescence Size Exclusion Chromatography
FX Fragment Exchange
GEF Guanine nucleotide Exchange Factor
GFP Green Fluorescent Protein
Gg *Gallus gallus*
GLP-1 glucagon-like peptide 1
GPCR G-protein coupled receptor
Gs *Galdieria sulphuraria*
HCAb heavy chain antibody
hGH Human Growth Hormone
HRP Horseradish Peroxidase
Hs *Homo sapiens*
Ig Immunoglobulin
IMAC Immobilised Metal Affinity Chromatography
ISR integrated stress response
KDELR KDEL Receptor
Kl *Kluyveromyces lactis*
LB Lysogeny Broth
LCP Lipidic Cubic Phase
LDAO lauryldimethylamine-N-oxide
LDS lithium dodecyl sulphate
LLG log-likelihood gain
LMNG 2,2-didecylpropane-1,3-bis-β-D-maltopyranoside
LSSR Local Structure Similarity Restraints
MFS major facilitator superfamily
Mm *Mus Musculus*
MMC methyl mercury chloride
MO monoolein
MR molecular replacement
mRNA messenger RNA
MST microscale thermophoresis
NA Numerical Aperture
nanoDSF nano differential scanning fluorimetry
NG n-nonyl-β-D-glucopyranoside
NM n-nonyl-β-D-maltoside
NMR Nuclear Magnetic Resonance
NT neurotensin
Ni-NTA nickel nitrilotriacetic acid
NTS1 neurotensin receptor 1

OD Optical Density
Os *Oryza sativa*
PBS phosphate buffered saline
PCR Polymerase Chain Reaction
PDB Protein Data Bank
PE pseudomonas exotoxin
PEG polyethylene glycol
PKA protein kinase A
PLP periodate lysine paraformaldehyde
QM quantum mechanic
qPCR quantitative PCR
RMSD root mean square deviation
Rn *Rattus norvegicus*
RNA ribonucleic acid
SAD single wavelength anomalous dispersion
Sc *Saccharomyces cerevisiae*
scFv single chain variable fragment
SD Standard Deviation
SDS-PAGE sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC Size Exclusion Chromatography
SEC-MALS Size Exclusion Chromatography - Multi-Angle Light Scattering
SFK Src family kinase
SGTA small glutamine-rich tetratricopeptide repeat-containing protein alpha
SNARE soluble N-ethylmaleimide-sensitive fusion protein attachment protein (SNAP) receptor
SOC Super Optimal Broth with Catabolite Repression
SPR Surface Plasmon Resonance
SR SRP receptor
SRP signal recognition particle
ssDNA single stranded DNA
SWEET Sugars Will Eventually be Exported Transporter
Syb37^{Cyto} Sybody37 - Cytoplasmic
Syb37^{Sec} Sybody37 - Secreted
Sybody synthetic nanobody
TB Terrific Broth
Tb *Trypanosoma brucei*
TCR T-cell receptor
TEV Tobacco Etch Virus
TFZ translation-function Z-score
THB triple helix bundle
TLS Translation/Libration/Screw
TM transmembrane
TMB 3,3′,5,5′-tetramethylbenzidine

TMD transmembrane domain
TMLA trimethyl lead acetate
TMRM tetramethylrhodamine-5-maleimide
TOG transporter-opsin-G protein-coupled receptor
UGGT UDP-glucose/glycoprotein glucosyl transferase
UPR unfolded protein response
VH variable domain of the heavy chain
VHH variable domain of heavy chain in HCAbs also referred to as Nanobodies
VL variable domain of the light chain
WT wild type
YPD Yeast Extract–Peptone–Dextrose

Supplementary Data

B

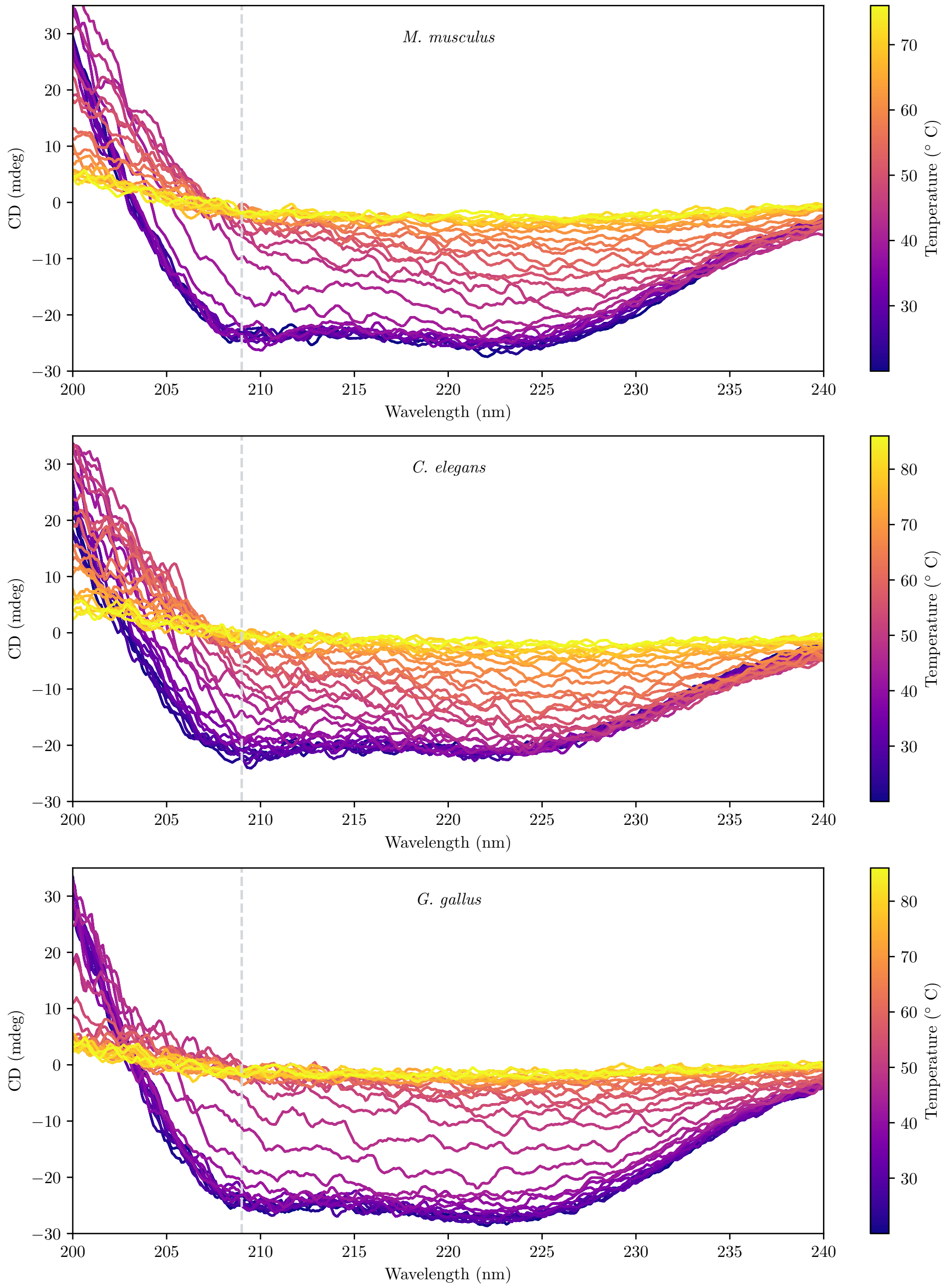


Figure B.1 Full CD melt profiles of *Mm*, *Ce* and*Gg* KDELR. *Mm* and *Gg* were measured in 20:1 DDM:CHS, while *Ce* was measured in 10:1. All measurements performed in 0.03% detergent with 10 mM potassium phosphate pH 5.0 and 50 mM sodium acetate.

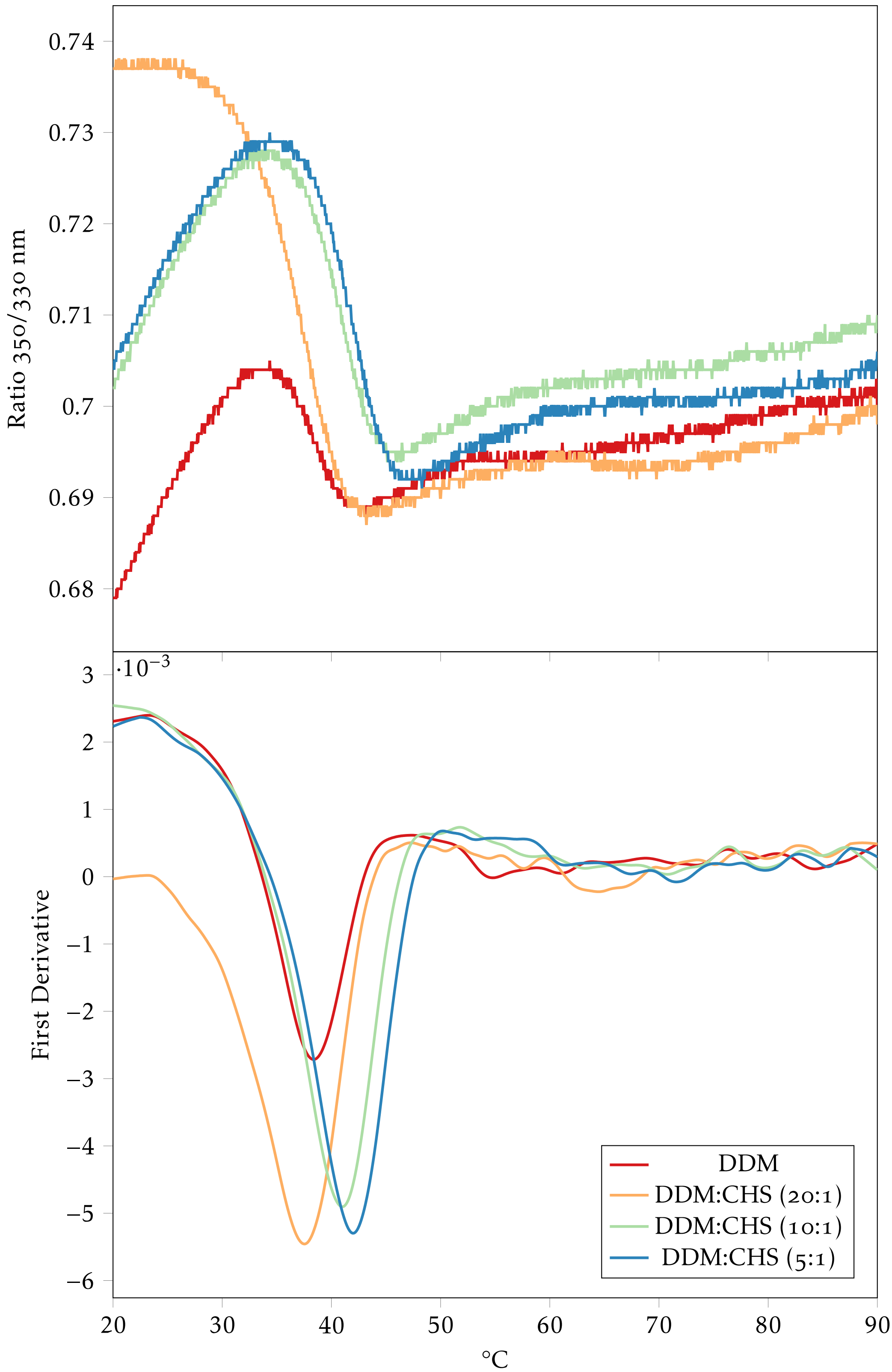


Figure B.2 *Mm* KDELR nanoDSF melts with increasing CHS. In 20 mM Tris pH 7.5, 150 mM NaCl, 0.03% detergent.

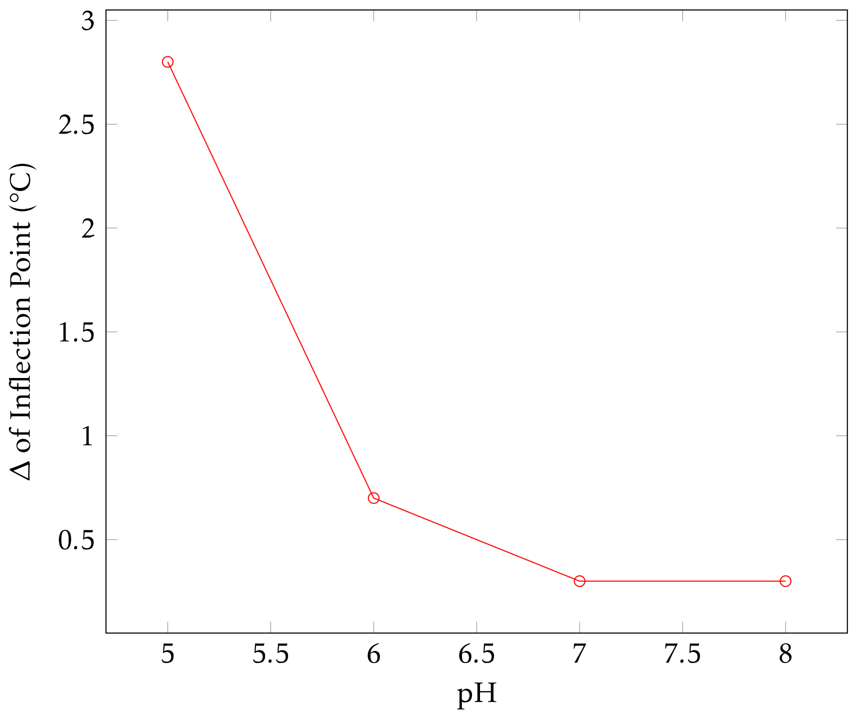


Figure B.3 pH-dependence of peptide stabilisation of *Mm* KDELR in nanoDSF. This plots the difference in inflection point for *Mm* KDELR (5 µM) with and without TAEKDEL peptide (20 µM). In 10 mM potassium phosphate at the indicated pH, 50 mM NaCl, 0.03% DDM 0.0015% CHS.

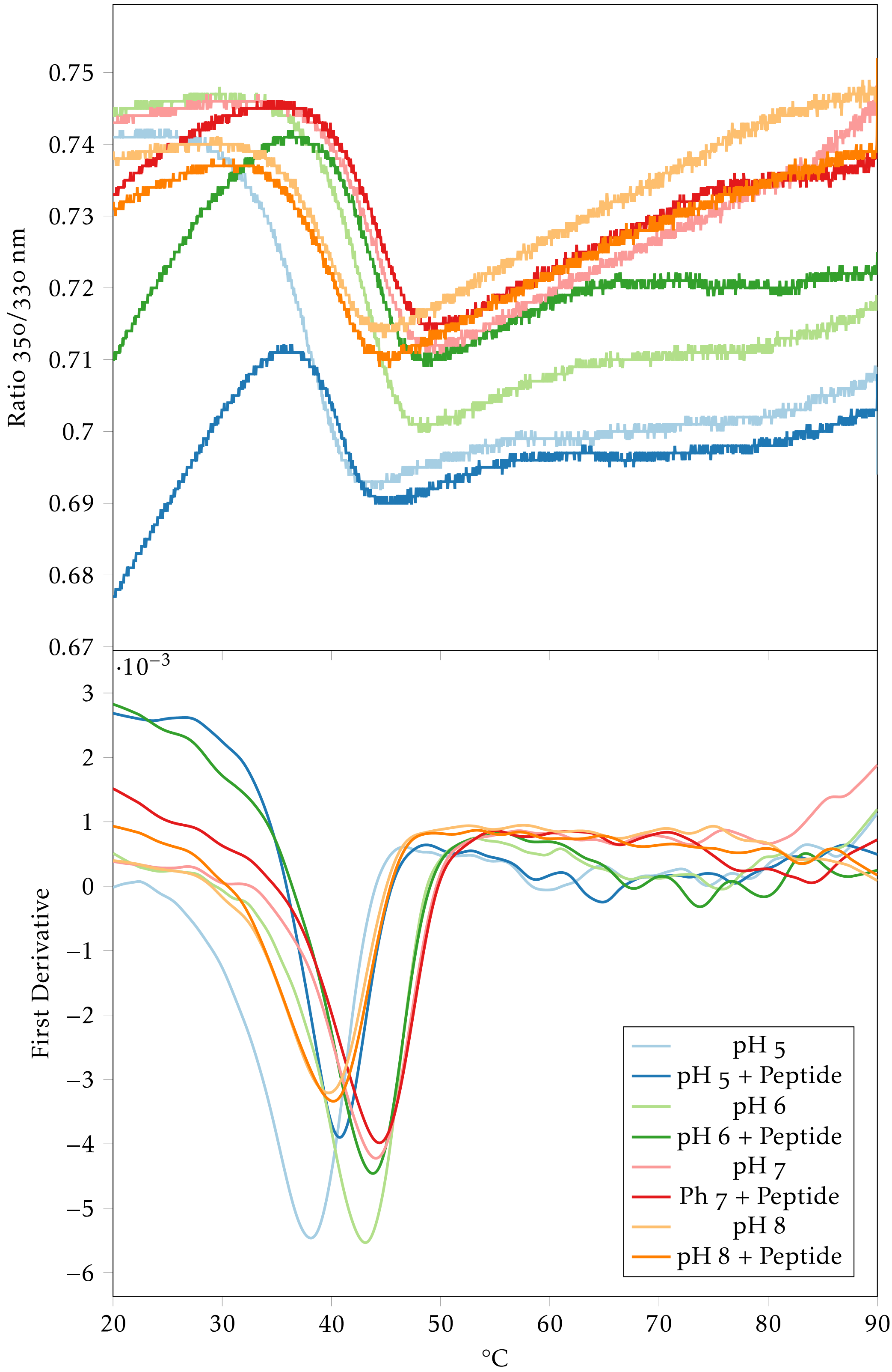


Figure B.4 Raw traces for *Mm* KDEL nanoDSF melts summarised in figure B.3.

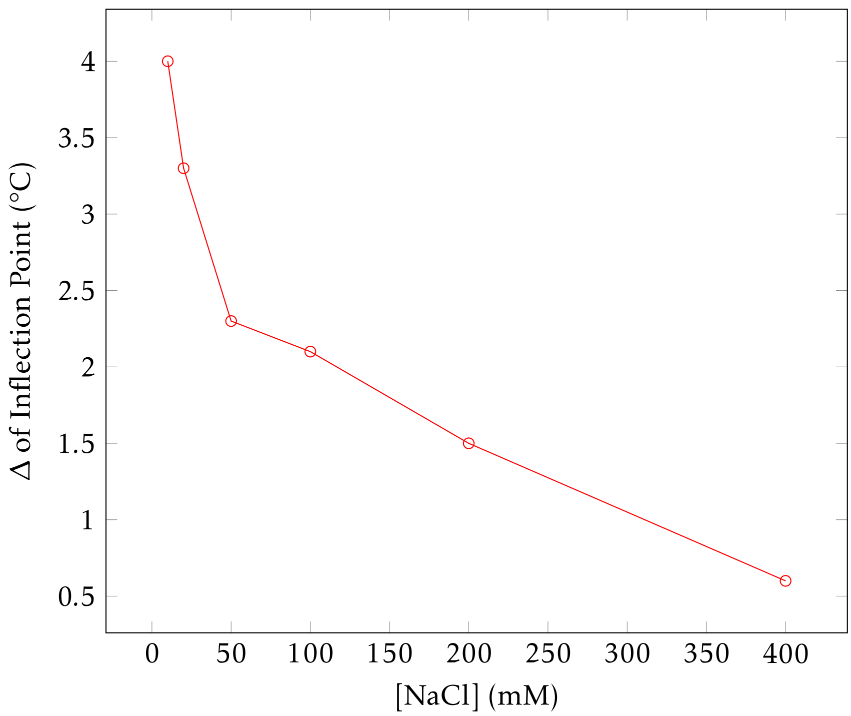


Figure B.5 Salt-dependence of peptide stabilisation of *Mm* KDELR in nanoDSF. This plots the difference in inflection point for *Mm* KDELR (5 µM) with and without TAEKDEL peptide (20 µM). In 10 mM potassium phosphate pH 5.0, variable NaCl, 0.03% DDM 0.0015% CHS.

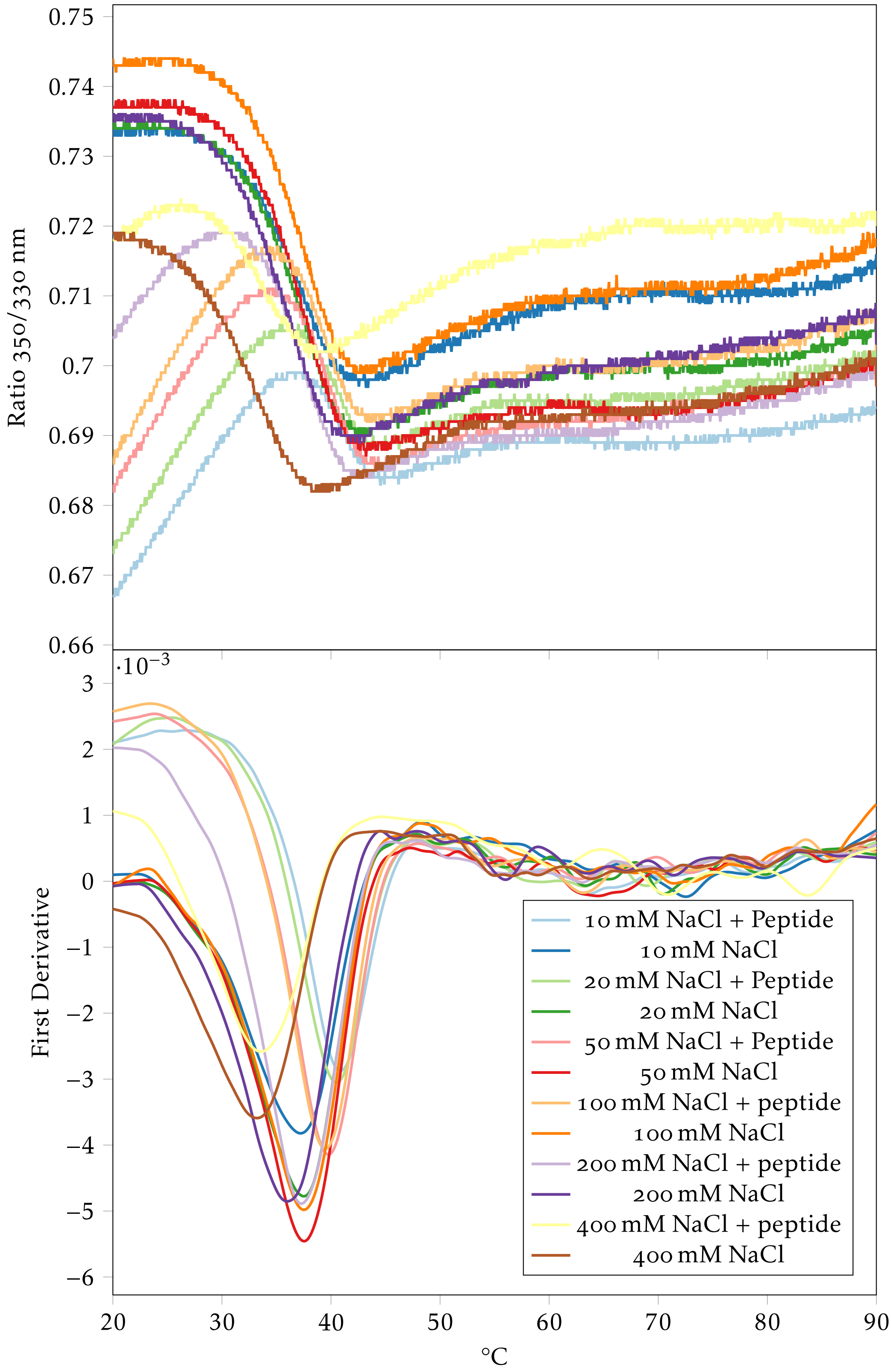


Figure B.6 Raw *Mm* KDEL nanoDSF melts summarised in figure B.5.

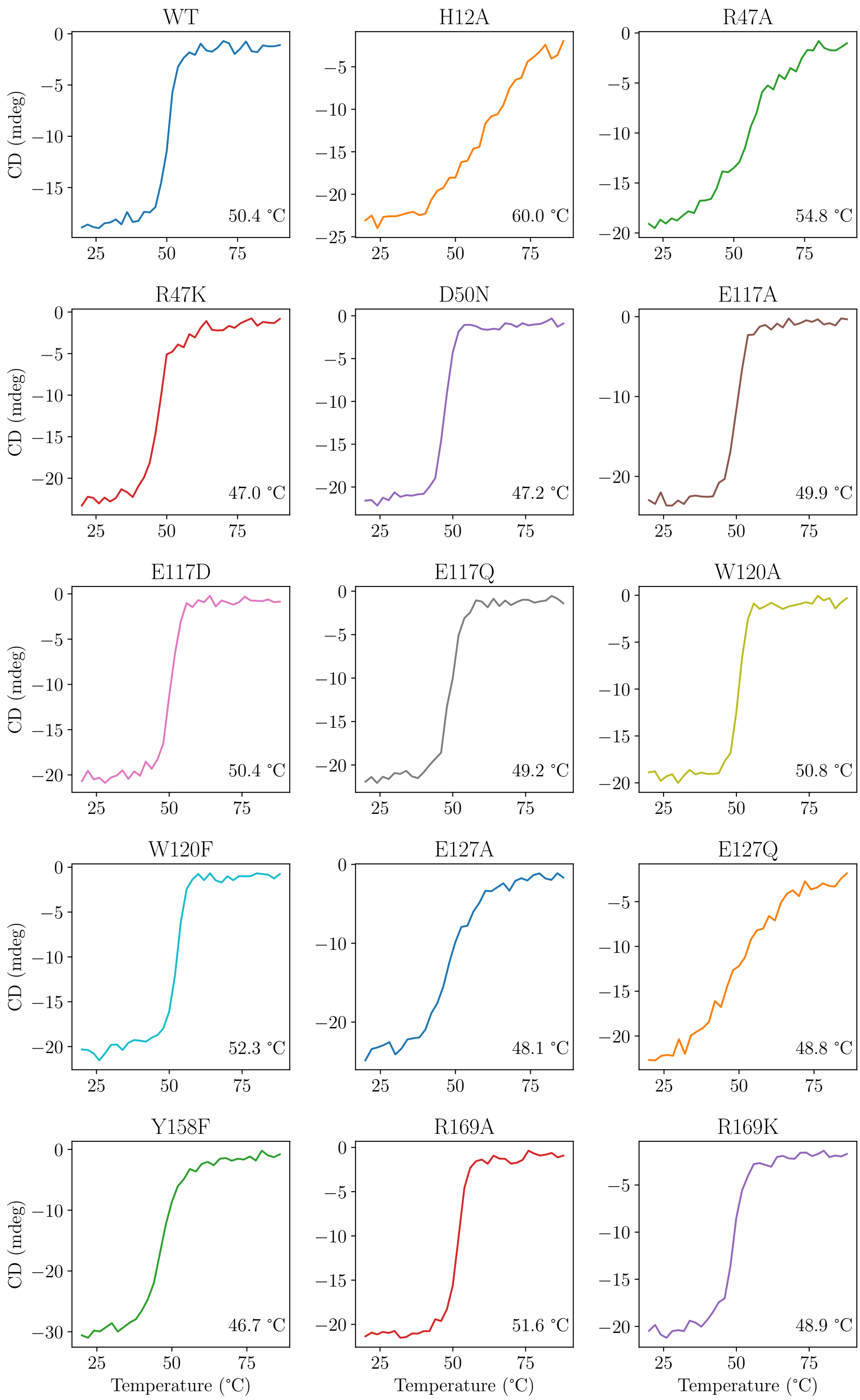


Figure B.7 CD melts profiles at 210 nm of the *Gg* KDELR and mutants with melting temperatures displayed in the bottom right, in 10 mM potassium phosphate pH 6.5, 50 mM Na_2SO_4 and 0.03% DDM 0.0015% CHS.

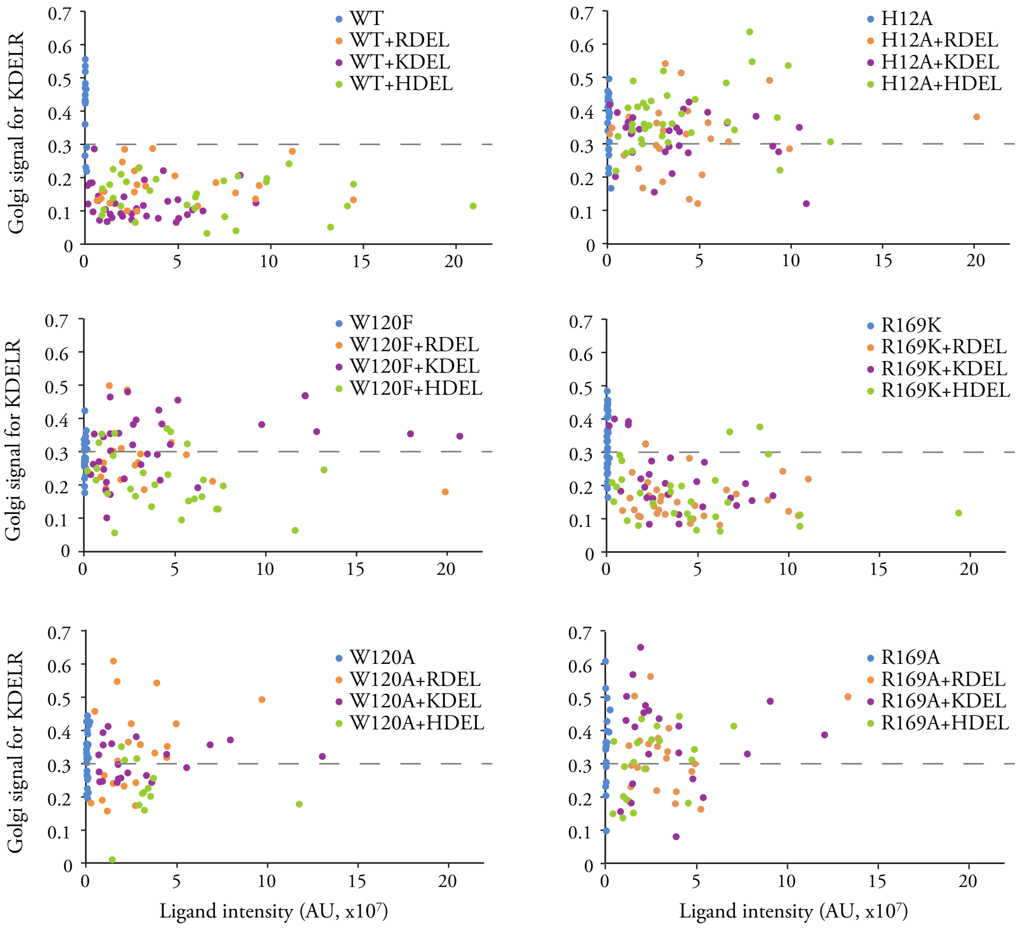


Figure B.8 Plots of Golgi signal for KDELR in response to K/H/RDEL ligands and the effect of ligand expression on the localisation of the KDELR. The dashed line indicates ER (below) or Golgi (above) localisation.

Vector Maps

C

C.1 Expression

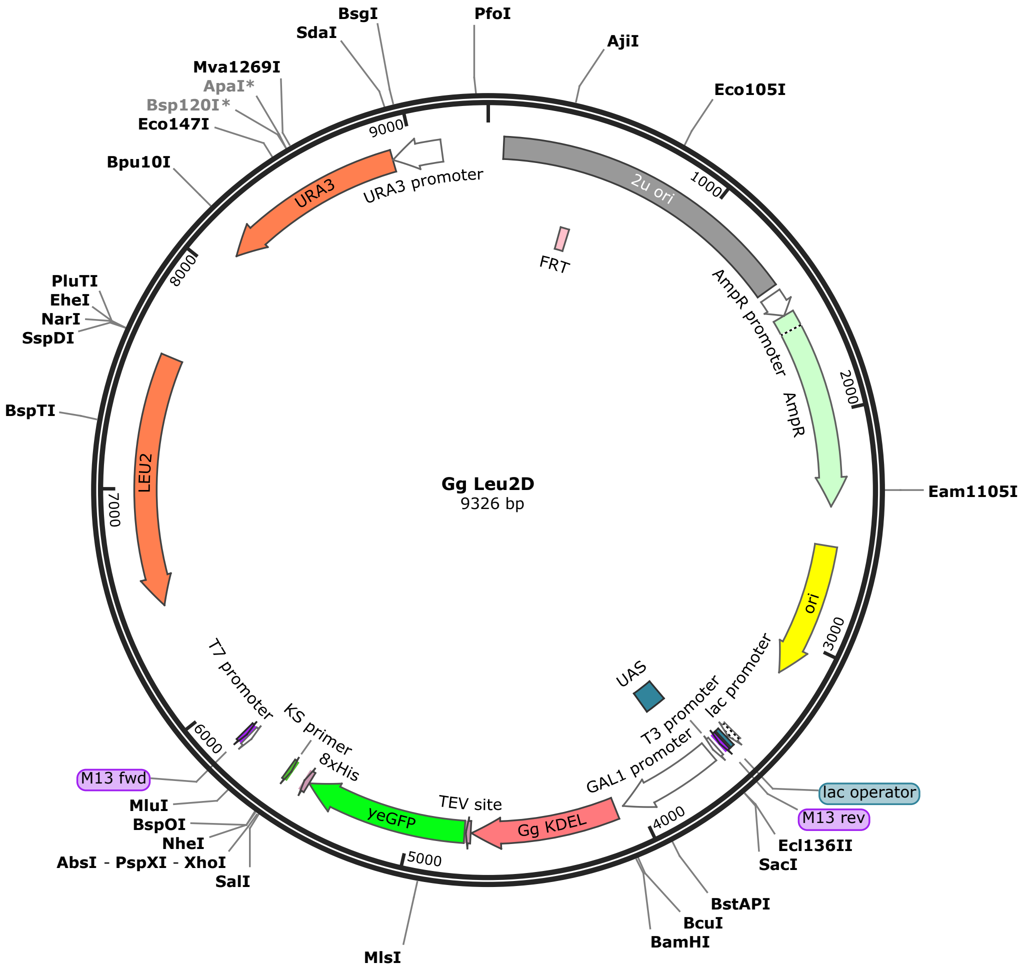


Figure C.1 Map of pDDGFP-Leu2D vector containing *Gg* KDELR, the same map applies to all homologues of the KDELR that were screened. This plasmid is available on Addgene: #123618.

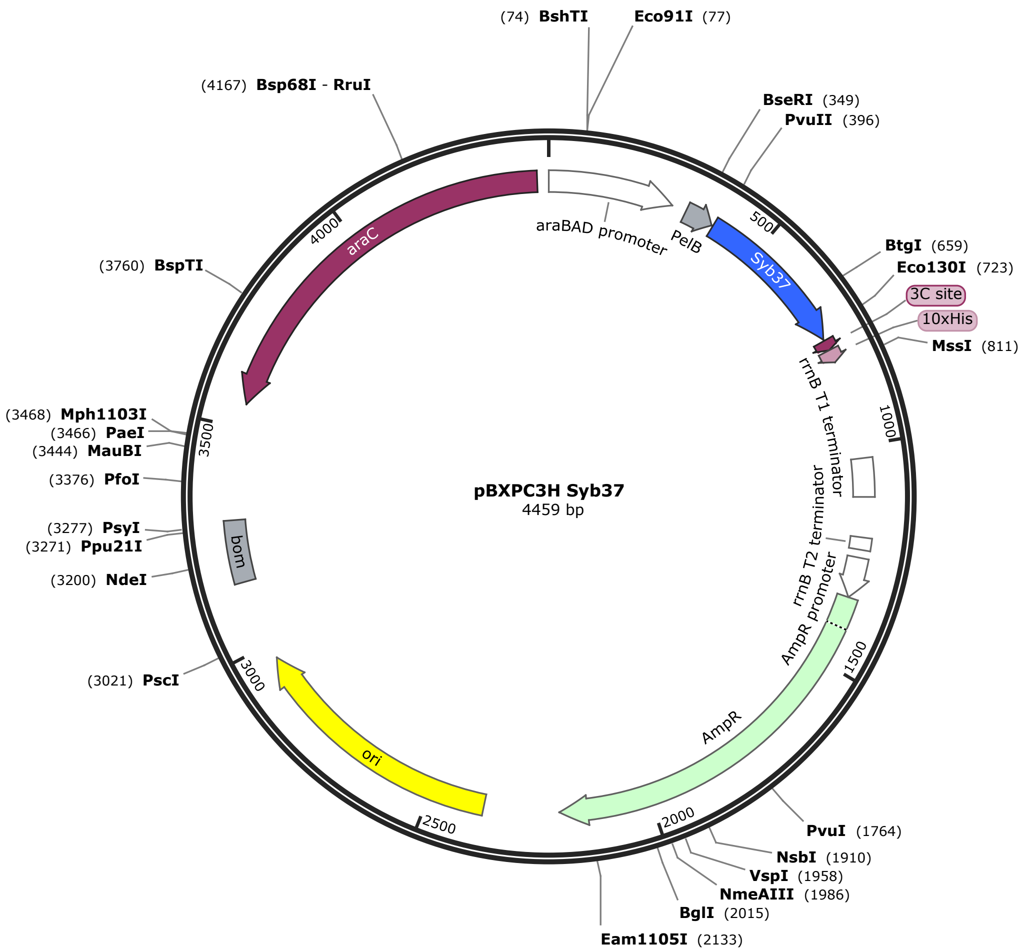


Figure C.2 Map of pBXPC3H vector containing Syb37. This plasmid is available on Addgene: #123627.

C.2 Sybody Pipeline

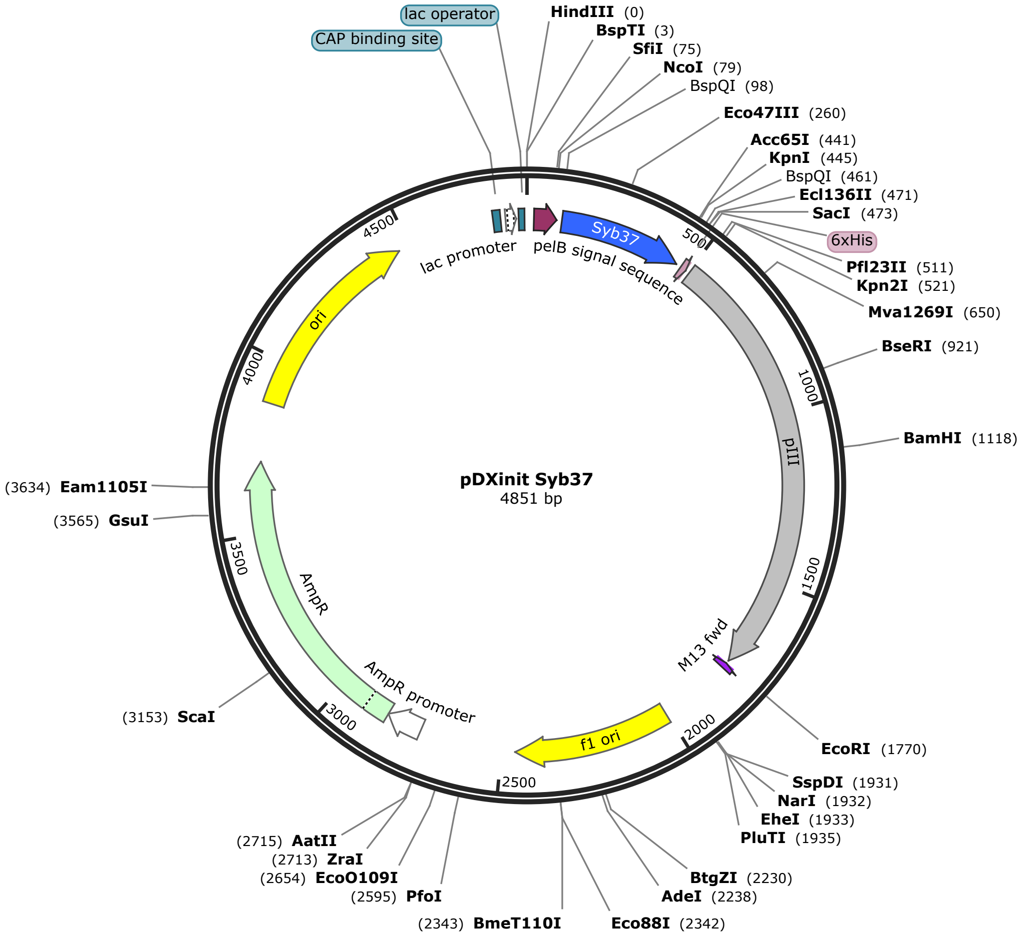


Figure C.3 Map of pDXinit containing Syb37. The empty plasmid is available on Addgene: #110101.

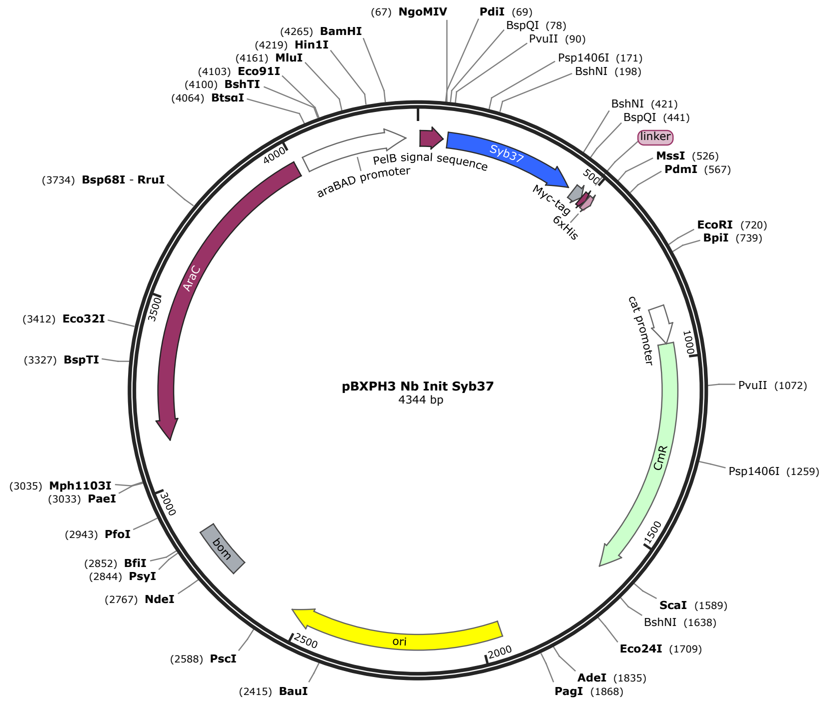


Figure C.4 Map of pBXPH3 Nb init vector containing Syb37. The empty plasmid is available on Addgene: #110100.

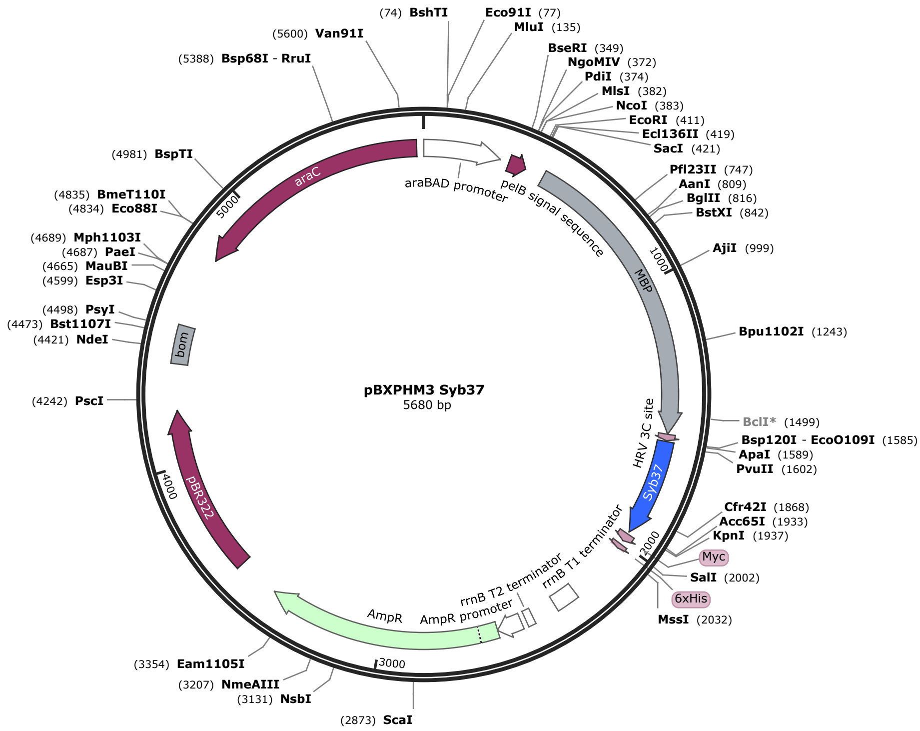


Figure C.5 Map of pBXPHM3 containing Syb37, this vector was used as an alternative espression vector for Sybodies that could not purified without an MBP tag. The empty plasmid is available on Addgene: #110099.

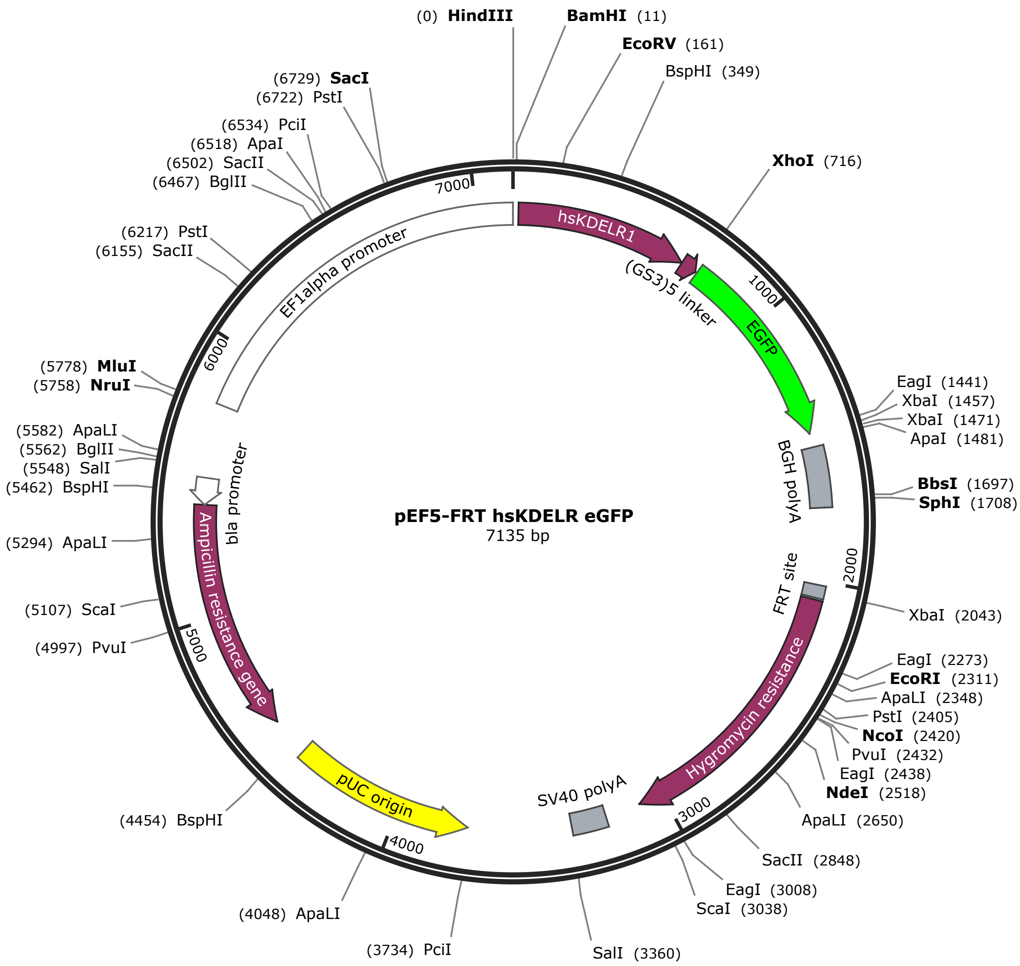


Figure C.6 Map of pEF5 TO FRT containing *Hs* KDELR1-GFP.

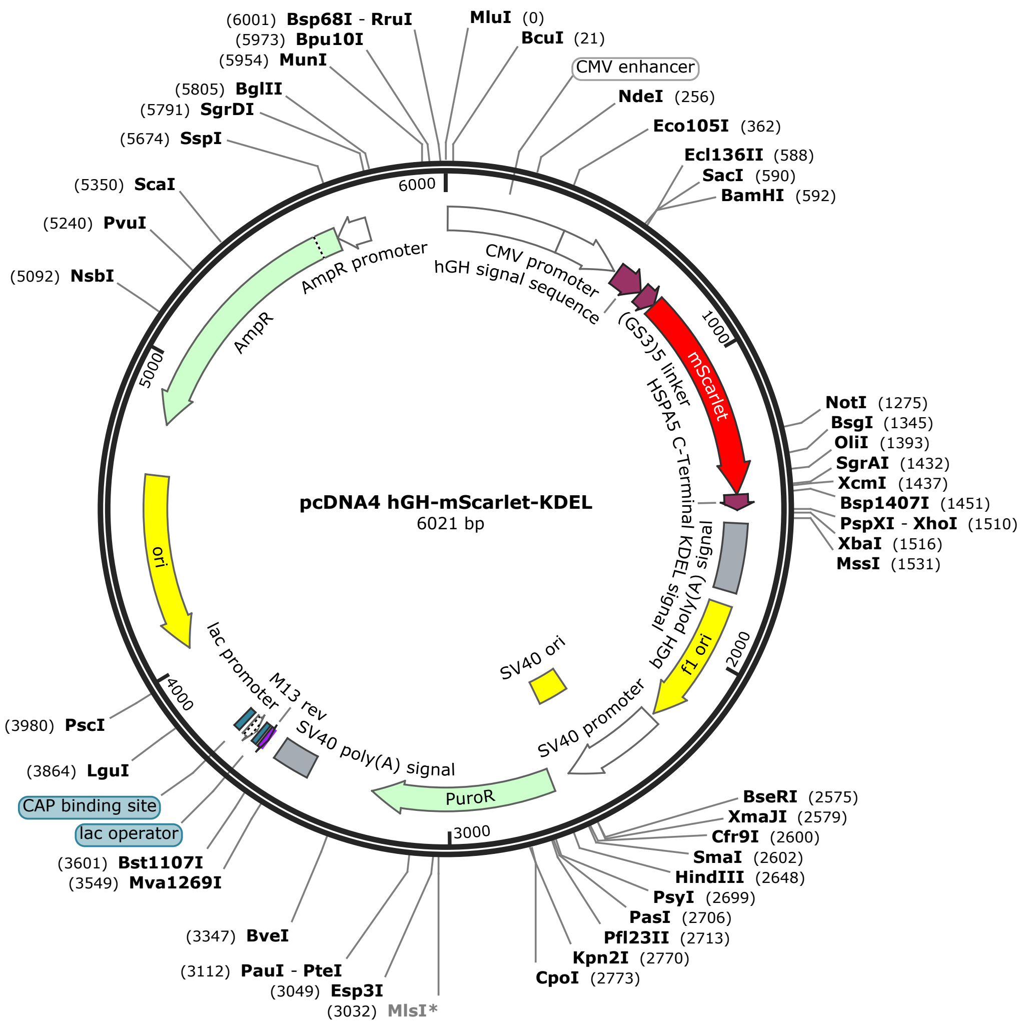


Figure C.7 Map of pcDNA4 containing hGH-mScarlet-KDEL.

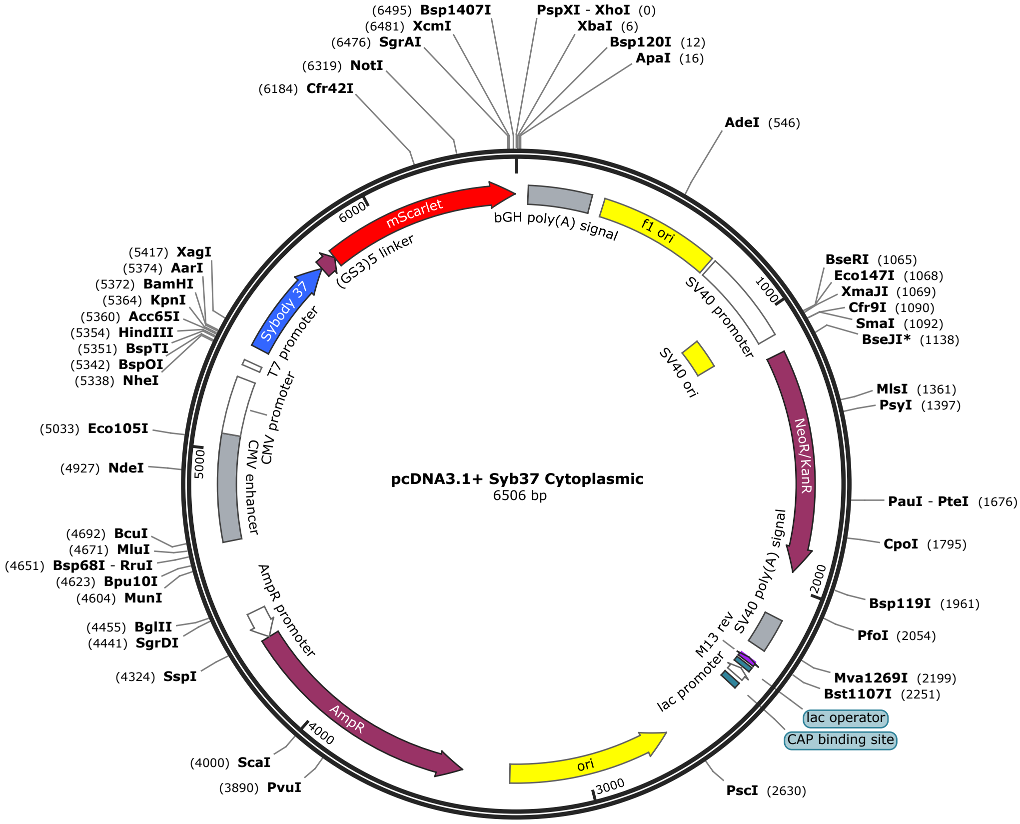


Figure C.8 Map of pcDNA3.1+ containing Syb37.

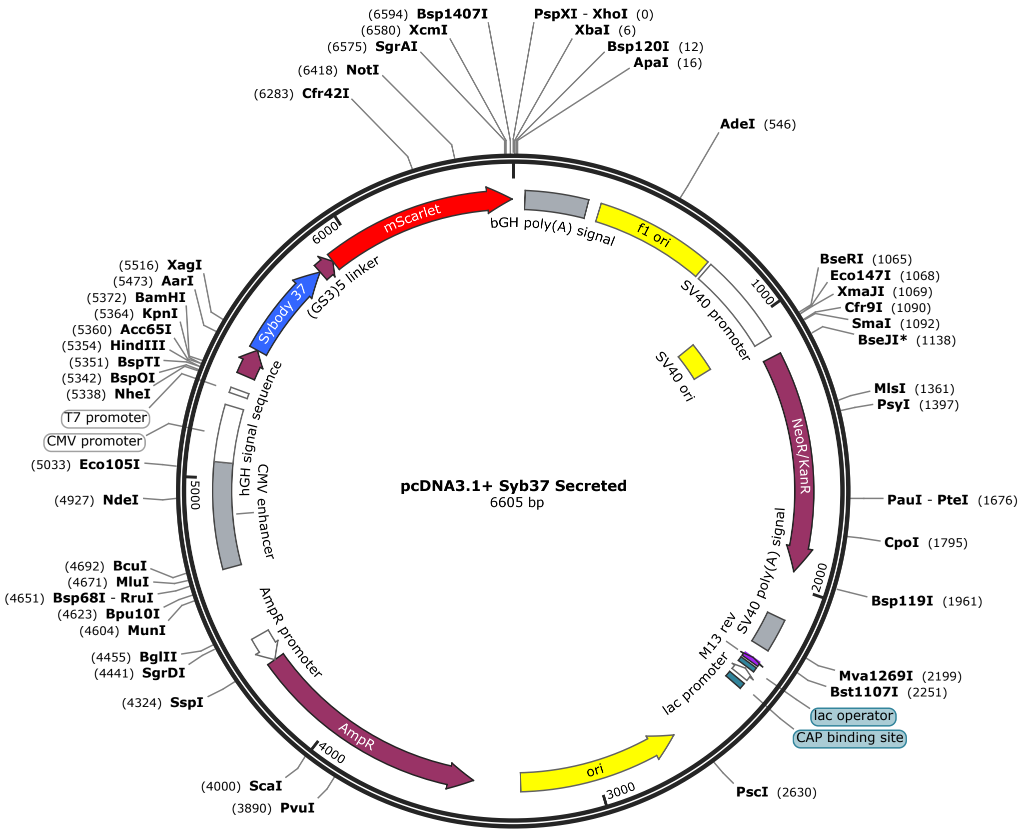


Figure C.9 Map of pcDNA3.1+ containing hGH-Syb37.

C.4 Mutagenesis

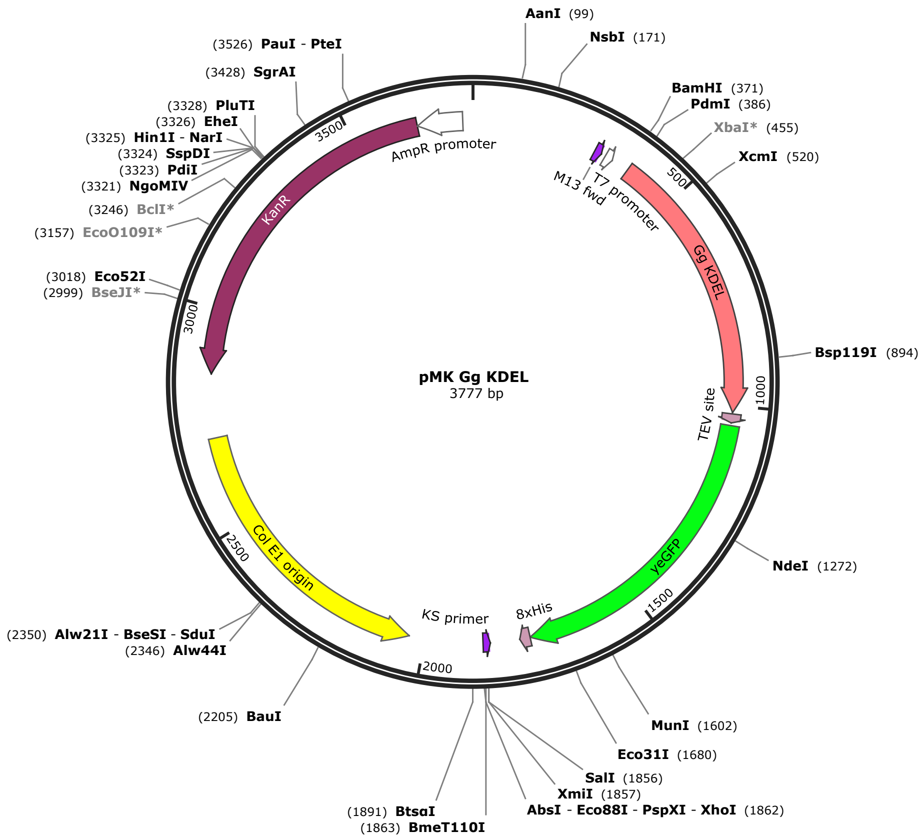


Figure C.10 Map of pMK containing *Gg* KDELR-GFP.

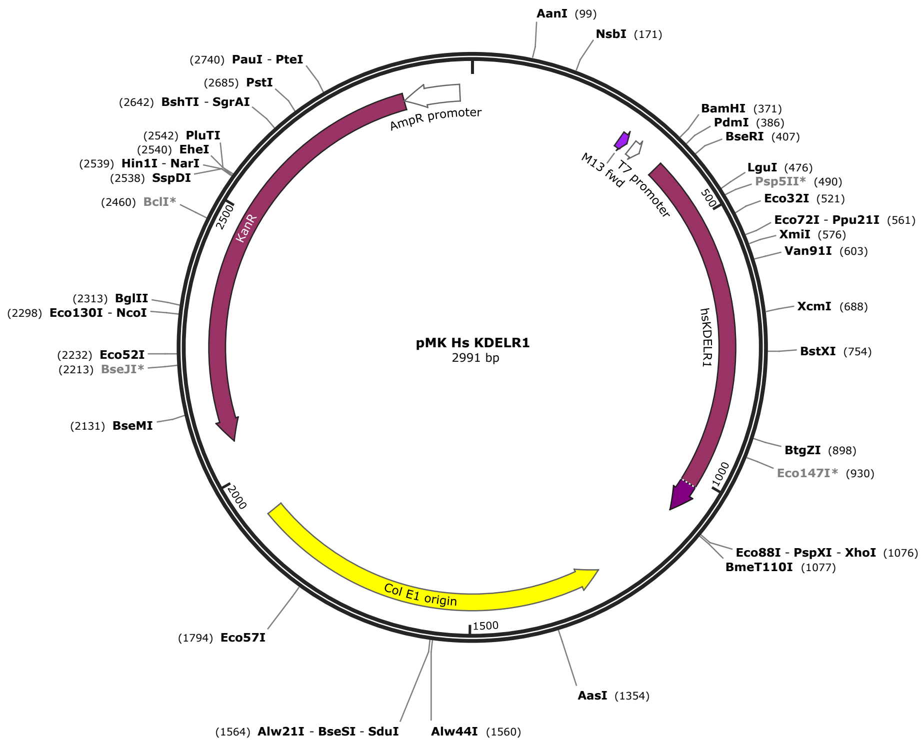


Figure C.11 Map of pMK containing *Hs* KDELR1-GS5.

Publications

D

Publications resulting from work in this thesis:

Bräuer P, Parker JL, Gerondopoulos A, Zimmermann I, Seeger MA, Barr FA, Newstead S. Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science*. 2019 03; 363(6431):1103. doi:10.1126/science.aaw2859.

Zimmermann I, Egloff P, Hutter CA, Kuhn BT, **Bräuer P**, Newstead S, Dawson RJP, Geertsma ER, Seeger MA. Generation of Synthetic Nanobodies against Delicate Proteins. *Nature Protocols*. 2019. In Press.

STRUCTURAL BIOLOGY

Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor

Philipp Bräuer[1], Joanne L. Parker[1], Andreas Gerondopoulos[1], Iwan Zimmermann[2], Markus A. Seeger[2], Francis A. Barr[1]*, Simon Newstead[1]*

Selective export and retrieval of proteins between the endoplasmic reticulum (ER) and Golgi apparatus is indispensable for eukaryotic cell function. An essential step in the retrieval of ER luminal proteins from the Golgi is the pH-dependent recognition of a carboxyl-terminal Lys-Asp-Glu-Leu (KDEL) signal by the KDEL receptor. Here, we present crystal structures of the chicken KDEL receptor in the apo ER state, KDEL-bound Golgi state, and in complex with an antagonistic synthetic nanobody (sybody). These structures show a transporter-like architecture that undergoes conformational changes upon KDEL binding and reveal a pH-dependent interaction network crucial for recognition of the carboxyl terminus of the KDEL signal. Complementary in vitro binding and in vivo cell localization data explain how these features create a pH-dependent retrieval system in the secretory pathway.

In eukaryotic cells, millimolar levels of chaperones required for protein folding in the endoplasmic reticulum (ER) are discriminated from newly synthesized secretory and membrane proteins which pass through the ER on their way to the Golgi apparatus (*1*). Luminal chaperones and ER-resident membrane proteins carry a C-terminal Lys-Asp-Glu-Leu (KDEL) sequence required for retention in the ER (*2*). When these proteins escape to the Golgi, they are recognized by an integral membrane protein, the KDEL receptor (KDELR), and retrieved back to the ER (*3*). The retrieval receptor ERD2 was discovered in 1990 as one of a number of ER retention defective (ERD) mutants in *Saccharomyces cerevisiae* (*3*–*5*). *ERD2* encodes the 26-kDa KDELR that enables yeast to retain ER luminal chaperones (*3*). It is predicted to have seven transmembrane domains and belongs to a large and diverse family of membrane proteins called the PQ-loop superfamily, which play important roles in lysosomal transport, tissue development, and nutrient transport in plants and bacteria (*6*). The KDELR is localized to the cis-Golgi, where it can efficiently capture escaped ER luminal proteins (*7*, *8*). The interaction with KDEL proteins is pH-sensitive, with maximal binding below pH 6 (*9*). The luminal pH of the cis-Golgi is 6.2, whereas that of the ER lumen is 7.2 to 7.4 (*10*, *11*), explaining how the KDELR could bind and release KDEL-carrying proteins in these organelles, respectively. Binding of a protein with a KDEL sequence to the receptor triggers incorporation of the complex into COPI vesicles, resulting in return of the complex to the ER (*12*). Once there, the complex dissociates and the receptor is rapidly trafficked back to the Golgi via COPII vesicles (*13*, *14*). The cytoplasmic surface of the KDELR is thought to mediate the interaction with the COPI vesicle machinery, although how this mediation occurs or how it is coupled to KDEL binding remains unknown (*13*, *14*). To elucidate the molecular basis for receptor-mediated retrieval in the secretory pathway, we determined the crystal structure of the KDELR in both the peptide-free and -bound states.

To determine the atomic structure of the KDELR, we screened several homologs from different eukaryotic sources to identify a receptor suitable for structural and biophysical characterization. *Gallus gallus* KDELR2 displays the characteristic pH-dependent binding for the KDEL peptide in the absence of any additional cofactors, with a dissociation constant of 1.25 ± 0.14 μM at pH 5.0 in detergent (Fig. 1A). Binding of the peptide is calcium independent (fig. S1), suggesting luminal calcium plays no role in the retrieval cycle. The structure was determined to 2.5-Å resolution at pH 9.0 in the peptide-free apo state (table S1). The structure has similarities to the eukaryotic SWEET transporter (*15*) (fig. S2), despite sharing only 24% sequence identity, and does not resemble the G protein–coupled receptor family of cell surface receptors. The receptor adopts a compact structure, consisting of seven transmembrane (TM) alpha helices arranged loosely in a hexagonal configuration when viewed from above (Fig. 1B). The first three N- and last three C-terminal helices form two internal triple helix bundles (THBs) arranged in a 1-3-2 sequence, connected with inverted topology by the linker helix TM4. This structural arrangement, whereby the THBs are inverted relative to each other by a pseudo twofold rotation axis in the plane of the membrane (fig. S3), suggests an evolutionary relationship to secondary active transporters (*6*).

The surface of the cytosolic face contains a prominent central band of negative charge running down its center (Fig. 1C). The negative charge is contributed by several acidic residues invariant in the mammalian KDELRs: Asp^{87}, Glu^{143}, Glu^{145}, and the C terminus of TM7 (figs. S4 and S5). Cell biological studies identified these and several additional residues in the cytoplasmic portion of human KDELR (*ERD2.1*) that result in retention within the ER when mutated (*16*). Many of these residues function to support the structural integrity of this electrostatic feature, suggesting this feature may form part of a diacidic COPII-binding ER-exit motif (*17*). The hydrophobic surface of the receptor is noticeably short, measuring 27 Å at its widest point (Fig. 1D), which is consistent with other membrane proteins resident in thin membrane bilayers of the ER and Golgi (*18*, *19*). The asymmetric position of this narrow hydrophobic belt suggests that the receptor projects outward from the cytosolic side of the membrane, exposing the negatively charged surface. Likewise, the receptor would be flush with the luminal side of the ER and Golgi membranes, where it recognizes KDEL-containing proteins. A large polar cavity is observed at the luminal side of the receptor, flanked by side chains from TMs 1 to 3 from the N-terminal THB and TMs 5 to 7 from the C-terminal THB, and measuring 13 Å by 15 Å by 12 Å (Fig. 1E). The electrostatic surface of the cavity is charged, with a pronounced dipolar character contributed by Arg^5 (TM1) and Arg^{169} (TM6), which are positioned opposite Glu^{117} (TM5) and Asp^{177} (TM7) (fig. S6). As discussed later, these residues form an integral part of the KDEL signal sequence recognition site.

Binding of the KDEL signal sequence to the human receptor is pH-dependent, yet the mechanism through which high-affinity binding depends on acidic pH is unclear (*9*, *20*). To address this question, we determined, to a resolution of 2.0 Å, a second structure of the *G. gallus* KDELR2 bound to the Thr-Ala-Glu-Lys-Asp-Glu-Leu (TAEKDEL) peptide at an acidic pH of 6.0 (table S1). The TAEKDEL peptide bound in the luminal-facing cavity adopts a vertical orientation with respect to the membrane, with the AEKDEL residues clearly resolved in the electron density map (Fig. 2A and fig. S7). Compared to the apo form, the peptide-bound receptor shows rearrangement of the side chains and helices forming the luminal-facing cavity (Fig. 2B and fig. S8). The cytoplasmic half of TM6 rotates inward, which results in Arg^{159} moving ~4.8 Å toward the peptide. This allows the C terminus of the KDEL ligand to be anchored in place through two salt bridge interactions to Arg^{159} (TM6) and Arg^{47} (TM2) (Fig. 2B). In this conformation, the luminal half of TM1 has moved outward to accommodate the peptide, with Arg^5 adopting a new rotomer configuration to interact with a carbonyl group on the peptide. Movement of TM1 also creates a pocket that accommodates the leucine side chain of the KDEL sequence. Recognition of the remaining side chains is predominantly mediated through electrostatic interactions to the sides of the

[1]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [2]Institute of Medical Microbiology, University of Zurich, 8006 Zurich, Switzerland.
*Corresponding author. Email: simon.newstead@bioch.ox.ac.uk (S.N.); francis.barr@bioch.ox.ac.uk (F.A.B.)

luminal-facing cavity. The positive amine group of the lysine interacts with Glu[117] (TM5) in a negatively charged pocket (Fig. 2C). The aspartate of the KDEL sequence makes a salt bridge to Arg[169] (TM6), whereas the glutamate interacts via a third salt bridge to Arg[5] (TM1) and makes a hydrogen bond interaction to Trp[166], which, similarly to Arg[159], moves inward to engage the peptide during the movement of TM6. The KDELR shows maximal binding between pH 5.0 and 5.4 with a gradual reduction in affinity until pH 7.0 (Fig. 1A). Our structures reveal that repositioning of TM6 upon peptide binding is stabilized through the formation of a short hydrogen bond, measuring ~2.5 Å, between Tyr[158] (TM6) and Glu[127] (TM5) (Fig. 2D and fig. S9). Formation of this short hydrogen bond would stabilize the new position of TM6, locking the peptide in place through its interaction with Arg[159]. Tyr[158] sits close to a strictly conserved histidine on TM1, His[12], forming an aromatic interaction. Two water molecules sit at the bottom of the peptide-binding site, coordinating the peptide carboxyl group to both Tyr[158] and His[12], which are further stabilized through a hydrogen bond to Asp[9] on TM1 (Fig. 2D and movie S1). Given its solvent-accessible position within the peptide-binding site, it is possible that protonation of His[12] facilitates the formation of the short hydrogen bond that stabilizes the repositioning of TM6, and thus acts as the pH sensor within the receptor. Supporting this proposal, conservative mutations in these residues resulted in the loss of KDEL binding in vitro (Fig. 3A). The loss of binding to the chicken receptor in vitro was matched by a failure of equivalent variants of the human receptor (ERD2.1) to respond to KDEL ligand overexpression in vivo, with the receptor remaining localized in the Golgi despite the presence of KDEL ligand (Fig. 3, B and C). Mutation of Arg[159] (TM6) resulted in a ligand-binding–defective protein trapped in the ER (Fig. 3, A to C), suggesting this residue is important for the conformational stability of the receptor.

In the peptide-bound structure, we also observed a pronounced effect at the cytoplasmic surface of the receptor, with TM7 moving away from TM5, ~14 Å relative to the apo receptor, and creating a new cavity facing the cytoplasm (Fig. 4, A and B). This movement results in the repositioning of a strictly conserved acidic residue, Asp[193] (TM7). Although an Asp[193]→Asn mutant in the human receptor was previously shown to bind KDEL peptide in vitro (*16*), our data show this variant remains trapped in the Golgi in vivo, even in the presence of an excess of KDEL ligand (Fig. 4, C and D). This supports the view that movement of TM7 and opening of the cytosol-facing cavity is important for ER retrieval via COPI. Consistent with this hypothesis, the movement of TM7 exposes a cluster of lysine residues (Lys[201], Lys[204], and Lys[206]) buried in the apo receptor (Fig. 4A) that could form an ER-retrieval motif (*21*, *22*). Additionally, the movement of TM6, discussed in relation to peptide binding above, causes the peptide chain linking TM5 and TM6 to shorten, as the length of TM6 is extended by one helical turn at the cytoplasmic end (fig. S10). This rearrangement results in the central band of negative charge, which was a prominent feature of the cytoplasmic face of the apo receptor, to lengthen and split into two equal regions (Fig. 4C). It is likely that such a drastic change in the electrostatics on the cytosolic-facing surface of the receptor plays a role in mediating the interaction between the KDELR and the COPI and COPII coatomer complexes during receptor trafficking. As already noted, the exposed lysine cluster on TM7 is similar to previously observed KKXX and KXKXX dilysine motifs, which are important in COPI-dependent Golgi-to-ER transport (*21*, *22*). To test their importance in the human KDELR, we mutated this motif and observed that the

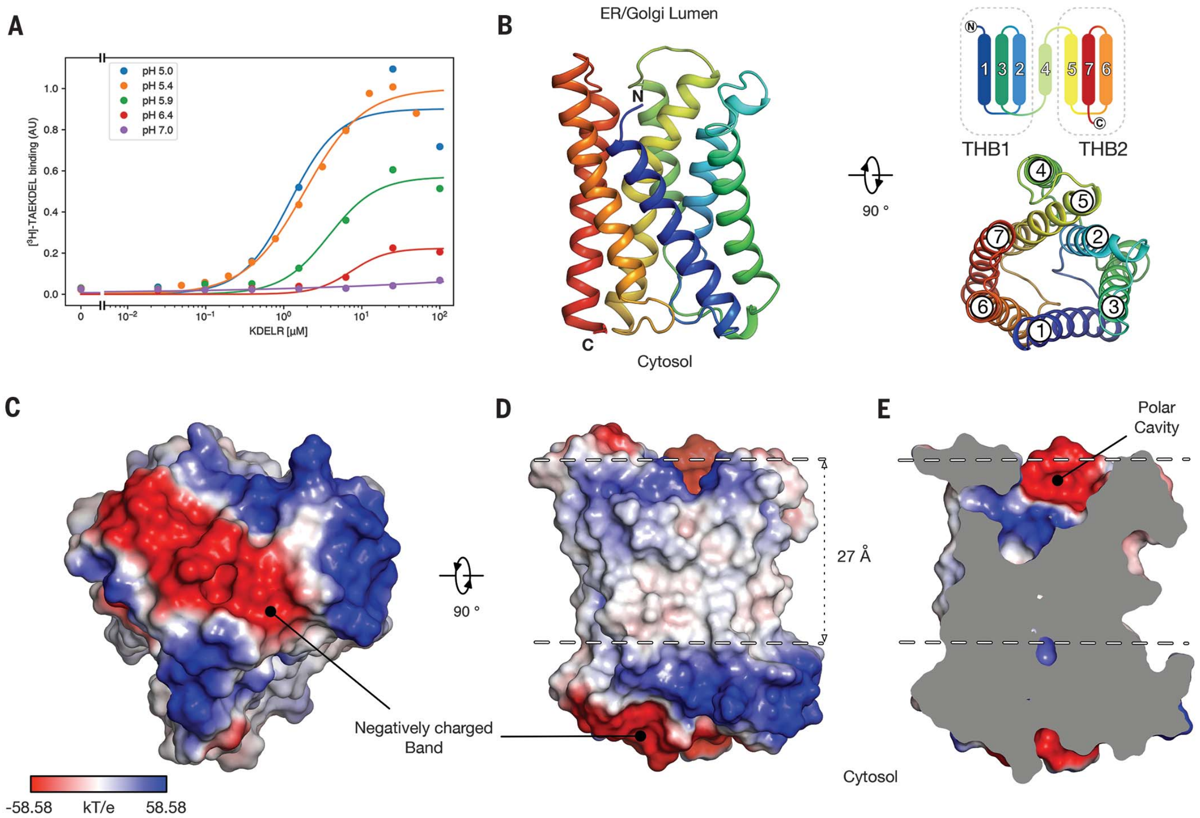


Fig. 1. Crystal structure of the KDELR. (**A**) Representative curves showing normalized binding of [^{3}H]-TAEKDEL peptide to the KDELR at different pH conditions. (**B**) Crystal structure of KDELR viewed from the Golgi membrane. The transmembrane helices are labeled by number. (Top right) Topology of the fold, highlighting the two triple helix bundles (THB1 and THB2). (**C**) Electrostatic surface representation of KDELR highlighting the negatively charged band on the cytosolic side of the receptor. (**D**) View in (A) rotated 90°, showing placement of the receptor with respect to the membrane bilayer (dashed lines). (**E**) Sliced-through volume representation of (D), highlighting the polar cavity on the luminal side of the receptor.

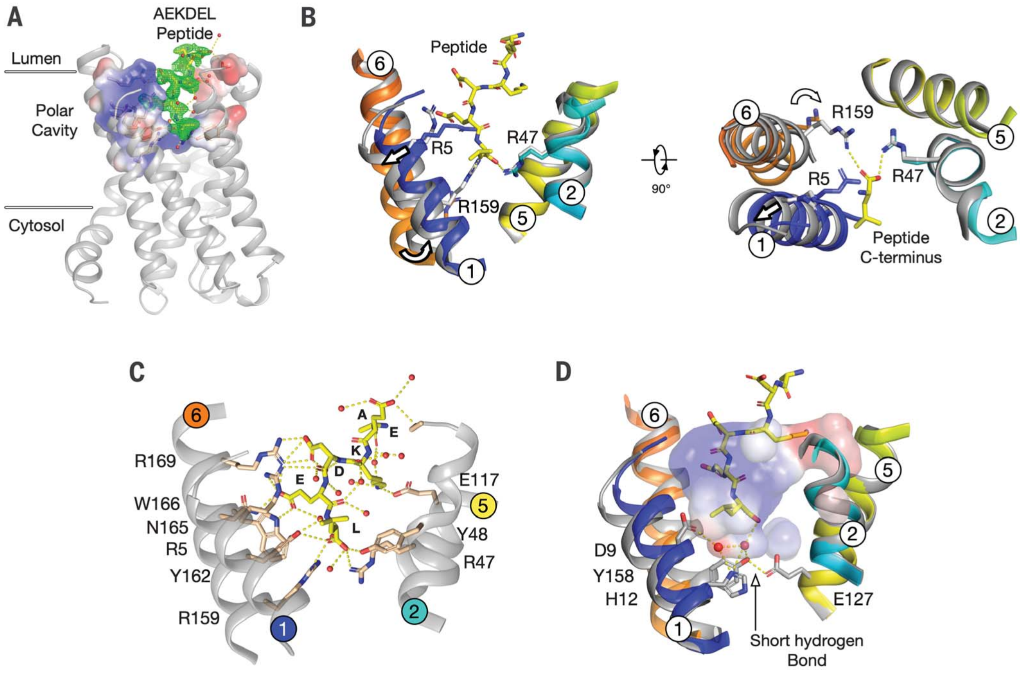


Fig. 2. Molecular basis for KDEL peptide recognition. (**A**) The polar cavity in KDELR is shown with contributing side chains represented as sticks and their electrostatic surface shown. The bound TAEKDEL peptide is shown with the *m*Fo-*D*Fc difference electron density (green mesh) used for model building displayed, contoured at 3σ. The N-terminal threonine was disordered in the maps and not modeled. Water molecules are represented as spheres (red) with hydrogen bonds as dashed lines (yellow). (**B**) Close-up view of the polar cavity showing the structural movement induced in TM1 and TM6 upon peptide binding. The apo structure is shown in colored helices, the peptide-bound structure in gray. In the top-down view (right), only the C-terminal leucine of the peptide is shown. (**C**) Close-up view of the polar cavity shown in (A). Water molecules are represented as spheres (red) with hydrogen bonds as dashed lines (yellow). (**D**) Overlay of the apo (color-coded helices) and peptide-bound structure (gray), highlighting the short hydrogen bond formed between Glu[127] and Tyr[158] at the base of the polar cavity after peptide binding. Two water molecules coordinate the interaction between the C terminus of the peptide with Tyr[158] and His[12]. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

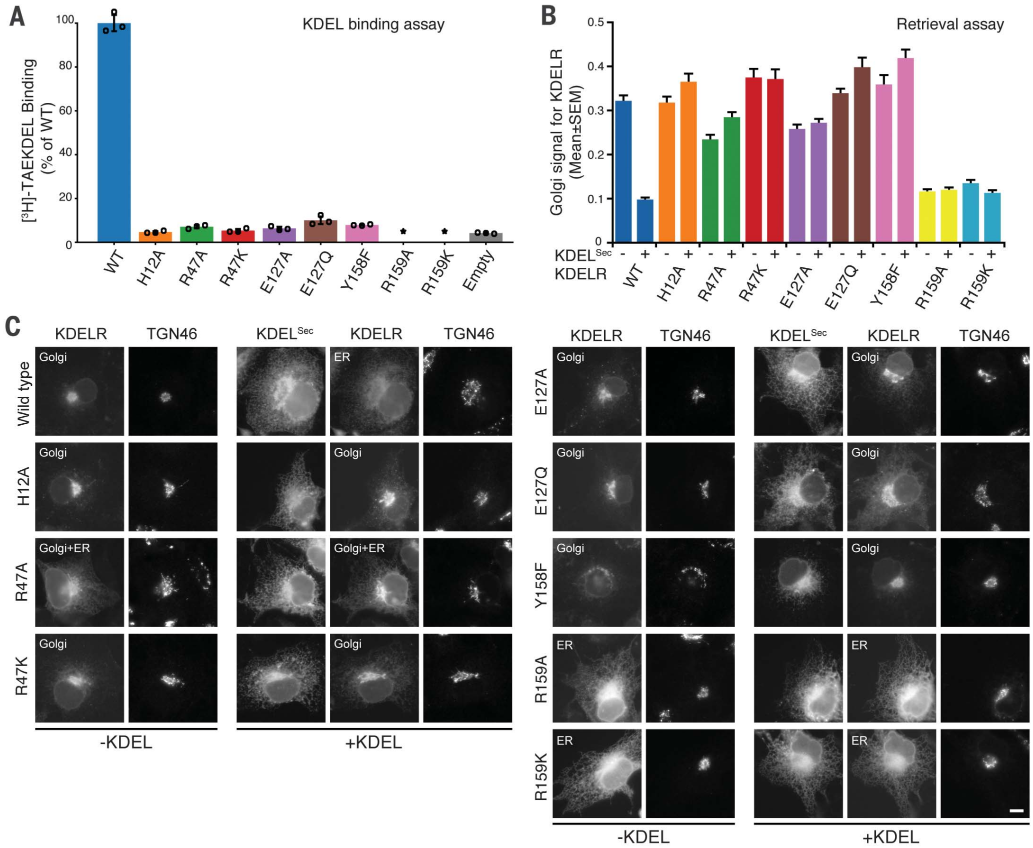


Fig. 3. Functional characterization of binding-site mutations. (**A**) In vitro binding assay using purified chicken KDELR. Asterisk indicates protein that could not be produced. (**B**) ER retrieval assays were performed in the absence (–) or presence (+) of KDEL^{Sec} and the Golgi signal for KDELR is plotted as mean ± SEM. (**C**) KDELRs were tested for KDEL^{Sec}-induced redistribution from Golgi to ER as in (B). TGN46 was used as a Golgi marker. Scale bar, 10 μm.

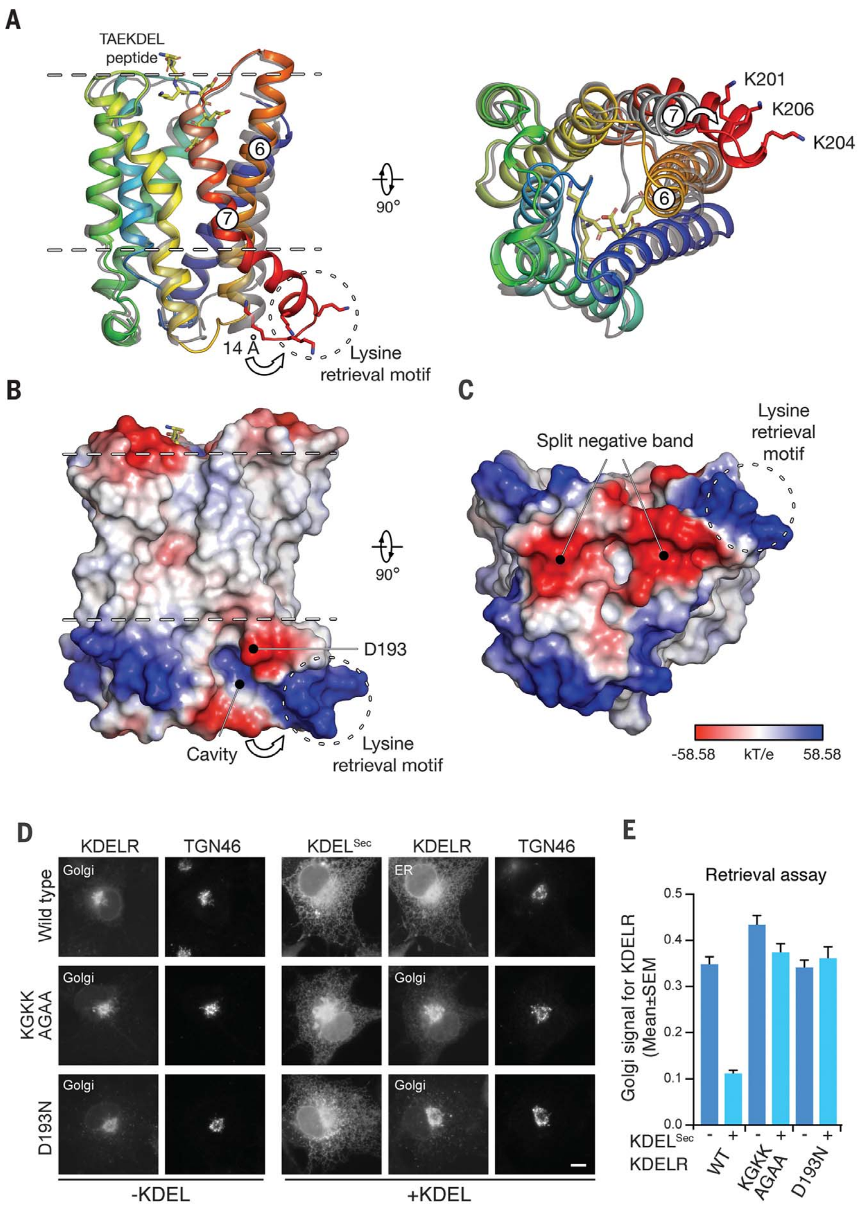


Fig. 4. Structural basis for exposure of COPI retrieval signal. (**A**) Overlay of the apo (gray) and peptide-bound (colored helices) crystal structures of the KDELR. The movement of TM7 is highlighted, exposing the C-terminal lysine side chains, indicated by the dashed circle. (Right) Cytoplasmic view of the receptor with the movement in TM7 indicated. (**B**) Electrostatic surface representation of (A). Dashed lines indicate rough position of the Golgi membrane, and the lysine retrieval motif is indicated by the dashed circle. After movement of TM7, a new cavity opens on the cytoplasmic side of the receptor, exposing Asp^{193}. (**C**) Cytoplasmic surface of the receptor, showing the substantial change in distribution of surface charge, most notably the disruption of the negative band present in the apo state. (**D**) Wild-type, Asp^{193}→Asn, and KGKK-dilysine motif mutant KDELRs were tested for KDEL^{Sec}-induced redistribution from Golgi to ER. TGN46 was used as a Golgi marker. Scale bar, 10 μm. (**E**) Golgi signal for KDELR in retrieval assays performed as in (D), in the absence (–) or presence (+) of KDEL^{Sec} is plotted as mean ± SEM.

protein remained localized in the Golgi upon KDEL ligand overexpression (Fig. 4, D and E). This motif therefore plays an important role during KDEL-mediated ER retrieval.

In the course of determining the crystal structure, we employed synthetic nanobodies, or sybodies, generated using an in vitro selection platform (*23*). A crystal structure, obtained at 2.23-Å resolution in the absence of KDEL peptide (table S1), demonstrates that the CDR3 loop of the sybody binds within the luminal-facing cavity yet fails to induce the movement of either TM6 or TM7 (fig. S11, A to E). Using a version of Syb37 targeted to the Golgi lumen, Syb37^{Sec}, we observed partial redistribution of KDELR-Syb37^{Sec} complexes from the Golgi to LAMP1-positive structures (fig. S11F). These results are consistent with the receptor no longer undergoing normal signal-mediated retrieval from the Golgi, but instead following the bulk flow pathway to the lysosome. Importantly, Syb37^{Cyto}, a cytoplasmic version unable to access the Golgi lumen, did not have this effect. The inability of the sybody to activate the KDELR either in vitro or in vivo highlights the importance of the specific interactions with the peptide in the luminal-facing cavity and subsequent movements in TM6 and TM7 that, we propose, define two conformations of the KDELR, which are important for anterograde and retrograde trafficking, respectively.

Taken together, our data support a mechanism whereby changes in the electrostatics on the cytoplasmic surface of the receptor, in combination with protonation of a key histidine and peptide binding, play an important role in retrieval of the receptor via the COPI pathway (fig. S12). A notable discovery from this work is the presence of mutually exclusive basic COPI and putative acidic patch COPII recognition motifs. TM7 plays a pivotal role in presenting these motifs. The structural change in TM7 is stabilized by an interaction network involving a conserved histidine located adjacent to the KDEL signal sequence binding pocket. This creates an elegant way for the KDELR to switch between the COPII ER-to-Golgi and COPI Golgi-to-ER trafficking pathways in a pH-dependent manner.

REFERENCES AND NOTES

1. L. Ellgaard, A. Helenius, *Nat. Rev. Mol. Cell Biol.* **4**, 181–191 (2003).
2. S. Munro, H. R. Pelham, *Cell* **48**, 899–907 (1987).
3. J. C. Semenza, K. G. Hardwick, N. Dean, H. R. Pelham, *Cell* **61**, 1349–1357 (1990).
4. F. M. Townsley, G. Frigerio, H. R. Pelham, *J. Cell Biol.* **127**, 21–28 (1994).
5. H. R. Pelham, K. G. Hardwick, M. J. Lewis, *EMBO J.* **7**, 1757–1762 (1988).
6. L. Feng, W. B. Frommer, *Trends Biochem. Sci.* **41**, 118–119 (2016).
7. M. J. Lewis, H. R. Pelham, *Cell* **68**, 353–364 (1992).
8. G. Griffiths *et al.*, *J. Cell Biol.* **127**, 1557–1574 (1994).
9. D. W. Wilson, M. J. Lewis, H. R. Pelham, *J. Biol. Chem.* **268**, 7465–7468 (1993).
10. M. M. Wu *et al.*, *Chem. Biol.* **7**, 197–209 (2000).
11. M. M. Wu *et al.*, *J. Biol. Chem.* **276**, 33027–33035 (2001).
12. I. Majoul, M. Straub, S. W. Hell, R. Duden, H. D. Söling, *Dev. Cell* **1**, 139–153 (2001).
13. S. R. Pfeffer, *Annu. Rev. Biochem.* **76**, 629–645 (2007).
14. N. Gomez-Navarro, E. Miller, *J. Cell Biol.* **215**, 769–778 (2016).
15. Y. Tao *et al.*, *Nature* **527**, 259–263 (2015).
16. F. M. Townsley, D. W. Wilson, H. R. Pelham, *EMBO J.* **12**, 2821–2829 (1993).
17. C. Barlowe, *Trends Cell Biol.* **13**, 295–300 (2003).
18. H. J. Sharpe, T. J. Stevens, S. Munro, *Cell* **142**, 158–169 (2010).
19. J. L. Parker, S. Newstead, *Nature* **551**, 521–524 (2017).
20. A. A. Scheel, H. R. Pelham, *J. Biol. Chem.* **273**, 2467–2472 (1998).
21. F. Letourneur *et al.*, *Cell* **79**, 1199–1207 (1994).
22. L. P. Jackson *et al.*, *Dev. Cell* **23**, 1255–1262 (2012).
23. I. Zimmermann *et al.*, *eLife* **7**, e34317 (2018).

ACKNOWLEDGMENTS

We thank the staff of beamlines I24 Diamond Light Source, UK, for assistance. **Funding:** This work was supported by Wellcome awards (102890/Z/13/Z, 109133/Z/15/A, and 097769/Z/11/Z) to S.N. and F.A.B. and SNSF Professorship of the Swiss National Science Foundation (PP00P3_144823) and Commission for Technology and Innovation CTI (16003.1 PFLS-LS) to M.A.S. **Author contributions:** S.N., J.L.P., M.A.S., and F.A.B. designed the experiments and interpreted the data. P.B., J.L.P., A.G., I.Z., and S.N. carried out the experiments and interpreted the data. P.B., J.L.P., F.A.B., and S.N. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Atomic coordinates for the models have been deposited in the Protein Data Bank under accession codes 6I6J (Syb37-KDELR2 complex), 6I6B (Apo structure), and 6I6H (peptide-bound structure). Expression plasmids containing the cloned genes detailed in this study are available from Addgene under a material agreement with the University of Oxford.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6431/1103/suppl/DC1
Materials and Methods
Figs. S1 to S14
Table S1
References (*24*–*35*)
Movie S1

5 December 2018; accepted 13 February 2019
10.1126/science.aaw2859

Science

Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor

Philipp Bräuer, Joanne L. Parker, Andreas Gerondopoulos, Iwan Zimmermann, Markus A. Seeger, Francis A. Barr and Simon Newstead

Science **363** (6431), 1103-1107.
DOI: 10.1126/science.aaw2859

Crystal structure of the KDEL receptor

Eukaryotic cells concentrate chaperones in the lumen of the endoplasmic reticulum (ER). These chaperones can be swept along the secretory pathway to the Golgi apparatus, from where they must be returned. For 20 years, cell biologists have known the identity of the KDEL (Lys-Asp-Glu-Leu) receptor responsible for this process, but the molecular basis for its function has remained elusive. Now, Bräuer *et al.* present crystal structures of the KDEL receptor, in both the apo ER state and KDEL retrieval signal –bound Golgi state. Comparisons of these two states identify the conformational switch that exposes the ER retrieval motif. The authors recapitulated the binding and release cycle of the receptor using purified components, confirming that the receptor is the minimal component required to bind KDEL ligands in the Golgi.

Science, this issue p. 1103

www.sciencemag.org/content/363/6431/1103/suppl/DC1

Supplementary Materials for

Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor

Philipp Bräuer, Joanne L. Parker, Andreas Gerondopoulos, Iwan Zimmermann, Markus A. Seeger, Francis A. Barr*, Simon Newstead*

*Corresponding author. Email: simon.newstead@bioch.ox.ac.uk (S.N.); francis.barr@bioch.ox.ac.uk (F.A.B.)

Published 8 March 2019, *Science* **363**, 1103 (2019)
DOI: 10.1126/science.aaw2859

This PDF file includes:

Materials and Methods
Figs. S1 to S14
Table S1
References

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/363/6431/1103/suppl/DC1)

Movie S1 (.mp4)

Materials and Methods

Cloning, expression and purification of KDELR

The genes encoding for *Gallus gallus* (*Gg*) KDELR2 (Uniprot: Q5ZKX9) and *Mus musculus* (*Mm*) KDELR1 (Uniprot: Q99JH8) were codon optimised and cloned into a modified form of a C-terminal GFP^{His} fusion vector for expression *Saccharomyces cerevisiae* strain BY5460 (*24*). Specifically, Wild-type and mutant forms of the KDELR gene were expressed by growing an overnight culture in synthetic complete medium minus leucine with 2% glucose (w/v). This culture was diluted tenfold in medium minus leucine with 2% lactate (v/v) in a 15 L fermentation vessel (Eppendorf BioFlo 415, DE); variants, as detailed in the text, were grown in flasks in 0.8 L of media. After 24 h, expression was induced through the addition of 1.5% galactose (w/v). After a further 24 h the yeast were collected and membranes prepared as described in (19). Wild type KDELR2 and variants were purified to homogeneity (Fig. S13), using standard immobilised metal-affinity chromatography (IMAC) in n-dodecyl-β-d-maltopyranoside (Glycon, DE) (DDM) with cholesteryl hemisuccinate (Sigma, UK) (DDM-CHS) detergent, as described previously (*25*). Briefly, membrane fractions were solubilised in 20 mM Tris pH 7.50, 0.1M NaCl, 1% DDM, 0.5% CHS for two hours. The soluble fraction was then bound to Ni-NTA resin (Pierce/Thermo, USA) in 20 mM Tris pH 7.50, 0.1M NaCl, 0.1% DDM, 0.05% CHS and subsequently eluted with 250 mM imidazole. After overnight dialysis with Tobacco Etch Virus (TEV) protease in 20 mM Tris pH 7.50, 0.15 M NaCl, 0.03% DDM, 0.0015% CHS, a reverse IMAC purification step was performed. The flow through was concentrated in a 10 kDa molecular weight cut off (MWCO) spin concentrator (VivaSpin, Sartorius) and applied to a Superdex 200 (10/300) column (GE Healthcare) in 20 mM Tris pH 7.50, 0.1M NaCl, 0.03% DDM, 0.0015% CHS.

Sybody selection and purification

Sybody selection was performed against C-terminally Avi-tagged and biotinylated *Mm* KDELR1, prepared using the same procedure as for *Gg* KDELR2. The protocols for sybody selection have been described elsewhere (23). A high affinity sybody (K_D 14.8 nM ± 1.7), Syb37, was identified from the loop library (23), which formed a stable complex with the receptor and co-eluted down a SRT-C SEC 300 column (Sepax, USA). Biolayer interferometry was performed on an Octet Red 384 (PALL, USA) using streptavidin biosensors which were loaded with biotinylated *Gg* KDELR2 at 0.1 mg mL^{-1} (Fig. S14). K_D measurements were performed using a serial dilution of the sybody from 500 nM to 7.8 nM. A 60 s baseline step in 20 mM Tris-HCl, 150 mM NaCl, 0.03% DDM, 0.0015% CHS was followed by a 300 s association step in the desired sybody concentration, followed by a 300 s dissociation step. Data were analysed in the Octet v9.0 software package. All raw data was baseline and reference subtracted, in-step corrected, y-axis aligned and filtered with a Savitzky-Golay filter.

Crystallisation and structure determination

Peptide-free Apo structure: The protein-laden mesophase was prepared by homogenizing monoolein (Sigma, UK) and 15.5 mg.mL^{-1} *Gg* KDELR2 in a 60:40 ratio by weight using a dual-syringe mixing device at 20 °C (*26*). Crystallisation was carried out at 19 °C using the Gryphon LCP robot in a 96 well glass sandwich plate (Thermo, USA) using 50 nL mesophase and 0.8 μL precipitant solution, consisting of 30% (v/v) PEG 500 DME, 100 mM Tris pH 9.0 and 100 mM Magnesium sulphate. Crystals were grown at 4 °C and harvested 10 days after setting up using a tungsten carbide glasscutter to open the plates and employing 30 – 75 μm micromounts (MiTeGen, USA) to harvest the crystals, which were subsequently cryocooled in liquid nitrogen and stored prior to data collection. Data from apo crystals were collected at the microfocus beamline I24 (Diamond Light Source, UK). All data were processed and scaled using the Xia2 (*27*) pipeline to DIALS (*28*) and AIMLESS (*29*) (Supplementary Table 1). No crystals were obtained for the apo receptor in acidic pH conditions. **Peptide bound structure:** *G*g KDELR2 was concentrated to 14.5 mg mL^{-1} and incubated with 6.4 mM TAEKDEL peptide (Cambridge Peptides, UK) on ice for one hour prior to crystallisation. Crystallisation plates were set up and left at 19 °C using precipitant 30% (v/v) PEG 600, 100 mM MES pH 6.0, 100 mM Sodium Nitrate. **Sybody-complex structure:** *Gg* KDELR2 Syb37 complex was concentrated to ~ 15 mg mL^{-1} after being bound on ice for 1 h and subsequently applied to an SRT-C SEC 300 SEC column (Sepax, USA). Fractions containing the complex were pooled and concentrated to 15.4 mg mL^{-1}. Crystallisation plates were set up at 19 °C as above, with crystals forming after 2 days in 30% (v/v) PEG 400, 100 mM Tris pH 9.0.

Phases were initially determined from the *Gg* KDELR2-Syb37 complex via molecular replacement using Phaser (*30*), employing PDB:5M13 as the search model. The initial phases produced a map that contained interpretable density for the receptor and a model was subsequently built into the density using O (*31*) and Coot (*32*), followed by refinement in Phenix (*33*) and BUSTER (*34*). The peptide-free and bound structures were phased using molecular replacement employing the KDELR model built into maps calculated from the

Syb37 complex crystals. The distance of the short hydrogen bond in the peptide bound complex (PDB: 6I6H) was fixed at 2.5 Å using the LINK command in BUSTER. However, refinement in Phenix suggested a shorter distance of ~ 2.3 Å.

Peptide binding assays

Binding assays were performed in 20 mM MES pH 5.4, 40 mM Sodium Chloride, 0.01% DDM 0.0005% CHS unless stated otherwise. 5 µL of ^{3}H-TAEKDEL (Cambridge Research Biochemicals, UK) at 20 nM was incubated with 5 µL of *Gg* KDELR, purified as above, or variants thereof at the desired concentration, at 20 °C for 10 min. The reaction was then filtered through a 0.22 µm mixed cellulose ester filters (Millipore, USA) using a vacuum manifold. Filters were then washed with 2 x 500 µL cold buffer. The amount of peptide remaining bound was measured using scintillation counting in Ultima Gold (Perkin Elmer). Experiments were performed at least three times, independently, to generate an overall mean and standard deviation (s.d). Data were normalised to the maximal binding at pH 5.4 and fit with a three-parameter non-linear regression model.

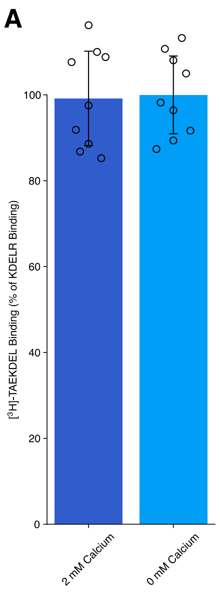
ER retrieval assays

Homo sapiens KDELR1 (Uniprot: P24390) was cloned into a pEF5/TO/FRT vector with a C-terminal 20 amino acid linker made up of 5 copies of Gly-Ser-Ser-Ser followed by GFP to create KDELR-GFP. Specific point mutations, described in the figures, were introduced using the Quickchange protocol (Stratagene, USA). To create the cytoplasmic targeted sybody ($Syb37^{Cyto}$) human codon optimised Syb37 produced using gene synthesis was cloned into a pcDNA3.1+ vector with the same C-terminal linker followed by mScarlet. The secretory pathway targeted Sybody37 construct ($Syb37^{Sec}$) was created by addition of amino acids 1-26 of human growth hormone (hGH) (the signal peptide) in-frame and upstream of the sybody start codon. To create the mScarlet-KDEL ligand construct, mScarlet with the N-terminal hGH signal peptide and the 16 C-terminal residues of human BiP at its C-terminus, containing the KDEL signal, was cloned into the pcDNA4 vector. COS7 cells were grown on 10 mm diameter 0.16-0.19 mm thick glass coverslips in DMEM containing 10% (v/v) bovine calf serum at 37°C and 5% CO_2. Cells were plated at 50,000 cells per well of a 6-well plate, each well containing 2 coverslips. For ER retrieval assays, the cells were transfected after 18h with 0.5 µg pEF5/TO/FRT KDELR-GFP and 0.5 µg pcDNA4 mScarlet-KDEL ligand (+ligand) or 0.5 µg pEF5/TO/FRT KDELR-GFP and 0.5 µg pcDNA4 empty vector (-ligand) diluted in 100µl

Optimem and 3 µl Mirus LT1 (Mirus Bio LLC). For Golgi localisation assays using Syb37, the cells were transfected after 18h with 0.5 µg pEF5/TO/FRT KDELR-GFP and 0.5 µg pcDNA3.1+ $Syb37^{Sec}$ or 0.5 µg pEF5/TO/FRT KDELR-GFP and 0.5 µg pcDNA3.1+ $Syb37^{Cyto}$. After a further 22 h, cells were washed twice with 2 mL of PBS, then fixed for 2 h in 2 mL PLP (2% w/v) paraformaldehyde in 87.5 mM lysine, 87.5 mM sodium phosphate pH 7.4, and 10 mM sodium periodate). Subsequently, coverslips were washed three times in 2 mL permeabilization solution 100 mM sodium phosphate pH 7.4, then permeabilised in 1 mg mL^{-1} BSA, 0.12 mg mL^{-1} saponin, and 100 mM sodium phosphate pH 7.4 for 30 min. Primary and secondary antibody staining was performed for 60 min in permeabilization solution at 22°C. Commercially available antibodies were used to detect TGN46 (sheep; AbD Serotec) and LAMP1 (mouse clone H4A3; BD). Coverslips were mounted on glass slides in Mowiol 4-88 and imaged with a 60×/1.35 NA oil immersion objective on an Olympus BX61 upright microscope (with filtersets for DAPI, GFP/Alexa-488, -555, -568, and -647 (Chroma Technology Corp.), a 2048x2048 pixel CMOS camera (PrimΣ; Photometrics), and MetaMorph 7.5 imaging software (Molecular Dynamics Inc.). Illumination was provided by a wLS LED illumination unit (QImaging). Image stacks of 3-5 planes with 0.3 µm spacing through the ER and Golgi were taken. The image stacks were then maximum intensity projected and the selected channels merged to create 24-bit RGB TIFF files in MetaMorph.

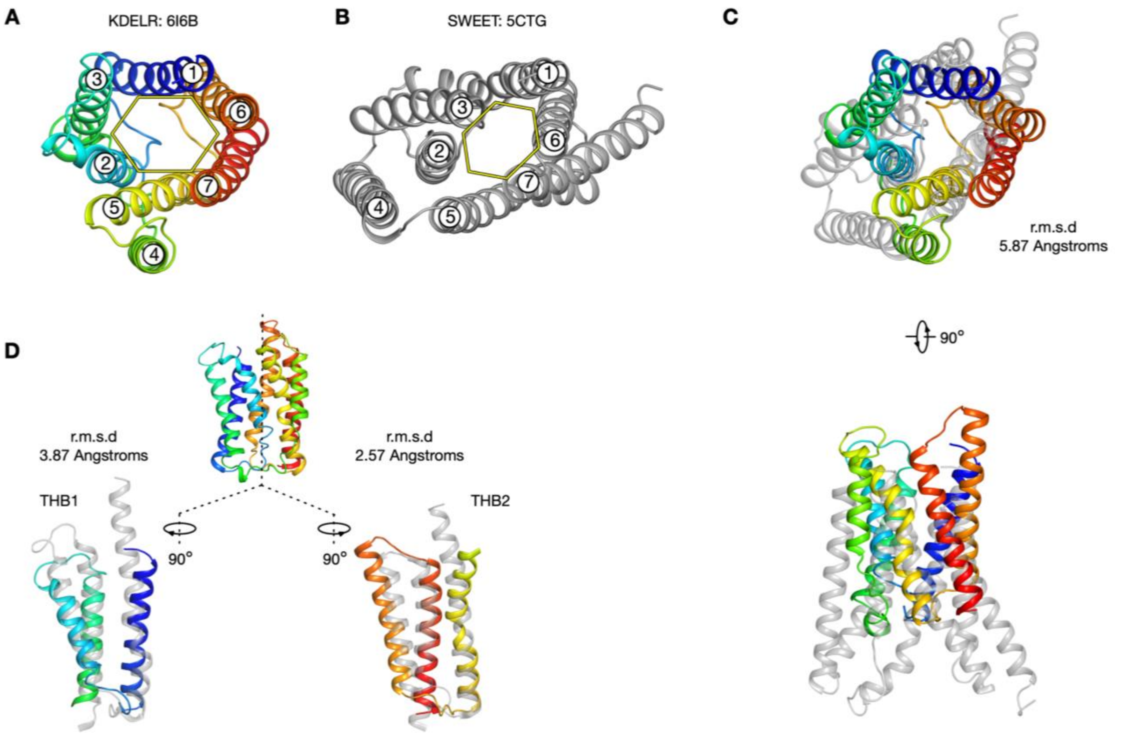
To produce the figures, images in 24-bit RGB format were cropped in Photoshop to show individual cells and then placed into Illustrator (Adobe Systems Inc., USA). To determine ER retrieval efficiency, the Golgi signal (integrated pixel intensity) for the KDEL receptor was measured in the region defined by the Golgi marker antibody in the presence (+) and absence (-) of KDEL ligand (*35*). Total cell KDELR signal (integrated pixel intensity) was also measured. For each cell, Golgi KDELR signal was divided by the total KDELR signal, to account for different expression levels, then plotted as a fraction in bar graph format. Error bars indicate the SEM. In Figure 3B n for WT=28, H12A=23, R47A=23, R47K=30, E127A=39, Y158F=22, R159A=23 and R159K=22. For Figure 4E n for WT=27, KGKK mutant=27 and D193N=44. Line intensity measurements of the different channels were used to examine the effect of Syb37 on KDELR localisation. These were plotted in line graphs.

Fig. S1.



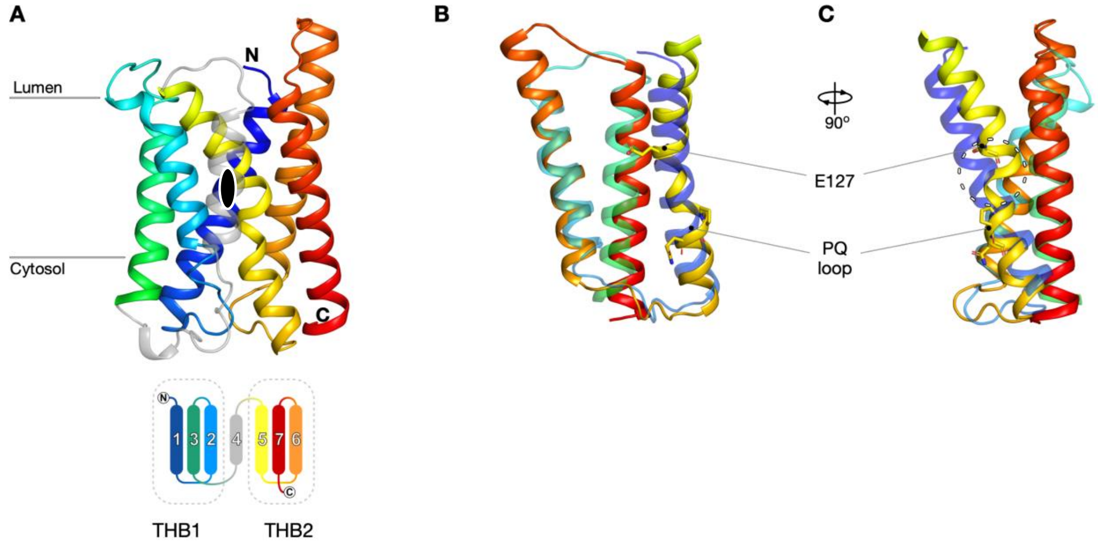
Peptide binding to the chicken KDEL receptor is independent of external calcium. A. Bar chart showing normalized binding of [^{3}H]-TAEKDEL peptide to the *Gg*KDELR in the presence (dark blue) and absence (light blue) of $CaCl_2$.

Fig. S2.



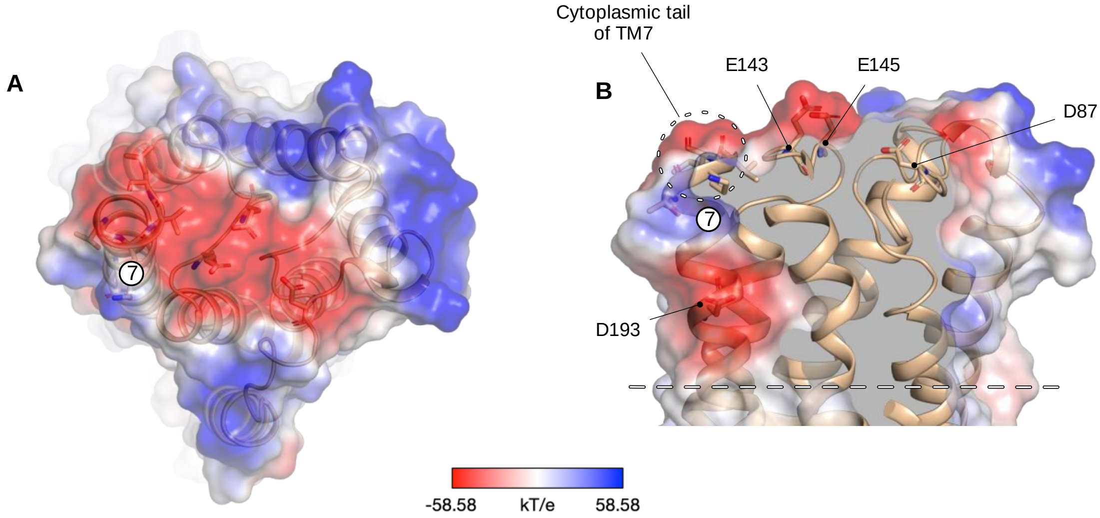
Structural comparison between the KDEL receptor and eukaryotic SWEET transporters. A. Luminal view of the KDEL receptor (PDB: 6I6B), coloured blue to red with TM helices labelled. The hexagonal arrangement of the TM helices is highlighted. **B.** Cytosolic view of the SWEET transporter from *O. sativa* (PDB: 5CTG), coloured grey, with TM helices labelled. A similar, though less obvious hexagonal arrangement of the helices can be discerned. **C.** Structural overlay of the KDEL receptor on the SWEET transporter using the cealign command in PyMOL. **D.** The KDEL receptor can be split into two triple helix bundles (THBs) as shown, which align better with the equivalent THBs in the SWEET transporter. Of particular note, TMs 1 and 5 are much shorter in the KDEL receptor relative to the vacuolar SWEET transporter (PDB:5CTG). TM4 was omitted for clarity.

Fig. S3.



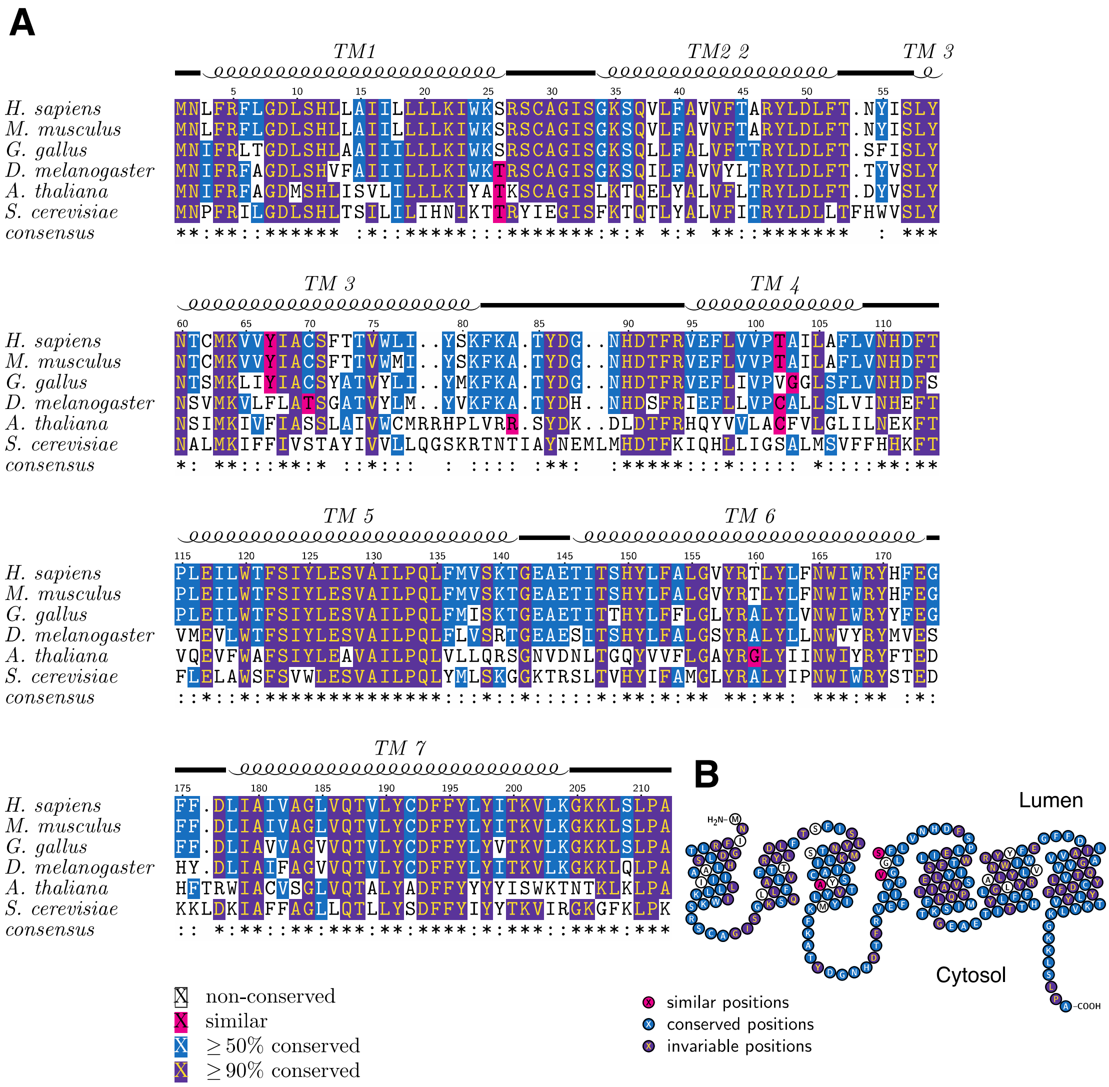
Structural analysis of the KDEL receptor. A. View of KDEL receptor in the plane of the membrane, colored red to blue. TM4, which acts as linker between THB1 and THB2, has been greyed out for clarity**.** Below, topology diagram indicating the arrangement of the helices. The pseudo-two fold rotation axis that relates THB1 with THB2 is shown as a black oval. **B.** THB1 can be superimposed onto THB2 with an r.m.s.d of 2.54 Å over 72 Cα atoms. The residues that form the 'PQ-loop', for which the superfamily is named, are located on TM5 and shown as sticks. The PQ-loop motif occurs close to Glu127, which is important in forming the short hydrogen bond that clamps the KDEL peptide in the receptor. **C.** TM6 adopts an unusual kink (dashed circle), which facilitates the correct positioning of Glu127 in the receptor and might explain the requirement for the PQ motif.

Fig. S4.



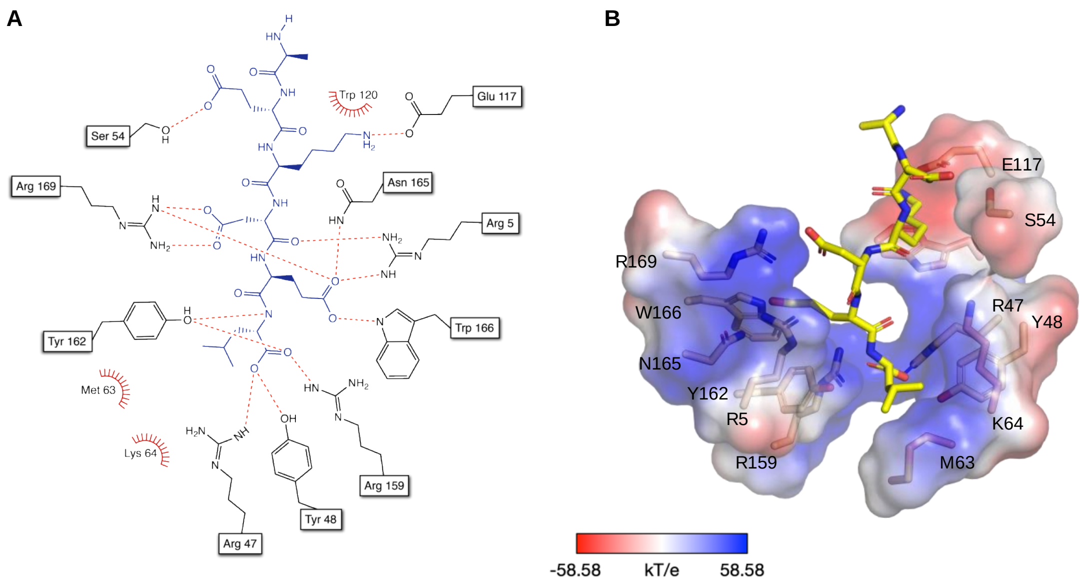
Analysis of the electrostatic surface charge on the KDEL receptor in the Apo state. A. Electrostatic surface of the cytosolic face of the KDEL receptor. Residues contributing to the prominent negatively charged patch are located on the loop regions connecting TM3 to TM4, and TM5 to TM6. The C-terminal end of TM7 also makes a considerable contribution to the negative charge, due to the free-carboxylic acid group and the carbonyl groups of the terminal residues. **B.** Sliced in view of the receptor in the membrane plane, cytosolic side facing up. Residues contributing to the negatively charged patch are labelled. D193 on TM7 is labelled, as this also contributes a distinctive negative charge close to the membrane surface (dashed line).

Fig. S5.



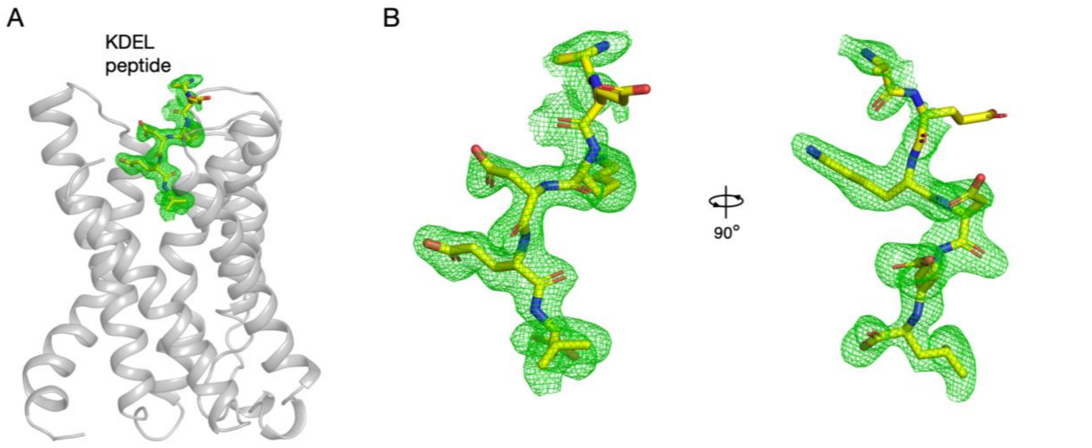
Sequence alignment and topology diagram of the KDEL receptor. **A.** Multiple sequence alignment between the KDEL receptors from *H. sapiens* (Uniprot: P24390), *M. musculus* (Q9CQM2), *G. gallus* (Q5ZKX9), *D. melanogaster* (O76767), A. *thaliana* (P35402*)*, *S. cerevisiae* (P18414). **B**. Topology diagram drawn using the crystal structure, colored coded for sequence conservation.

Fig. S6.



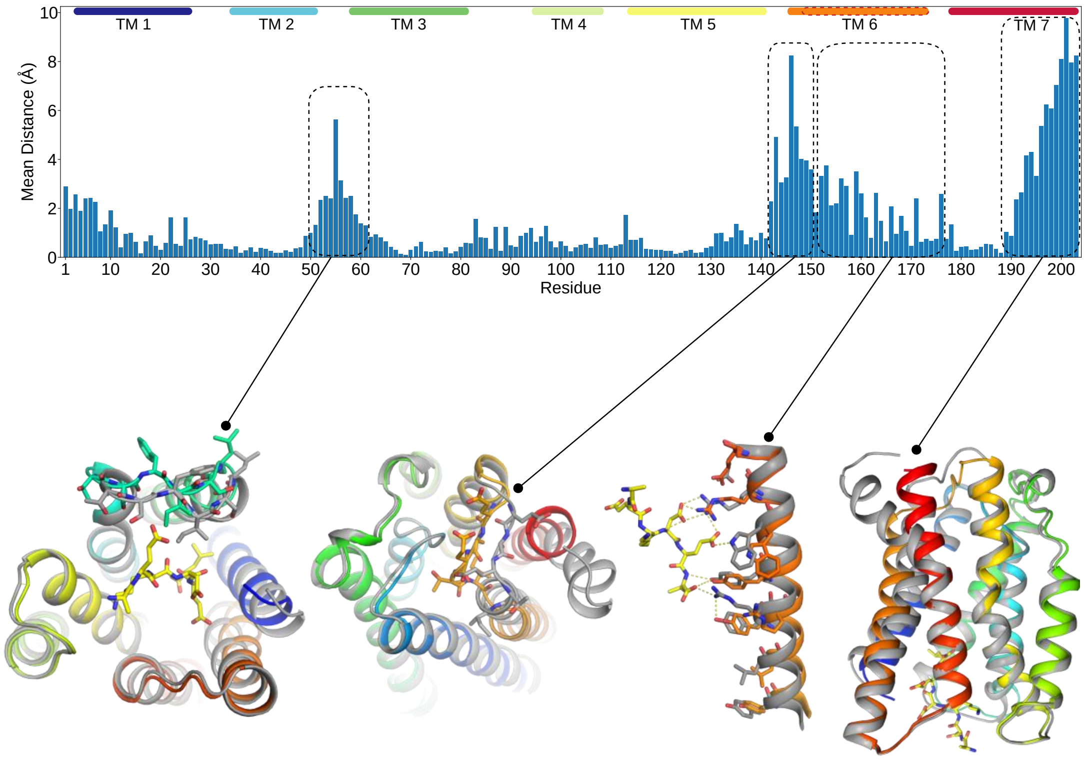
Molecular basis for KDEL peptide recognition. A. Schematic showing the interactions made between the KDEL peptide and receptor. Only the EKDEL region of the TAEKDEL peptide is shown as the two N-terminal residues do not interact with the receptor. Hydrogen bonds are shown as dashed lines (red), hydrophobic interactions as semi-circles (red). **B.** Electrostatic surface of the peptide binding cavity depicted**.**

Fig. S7.



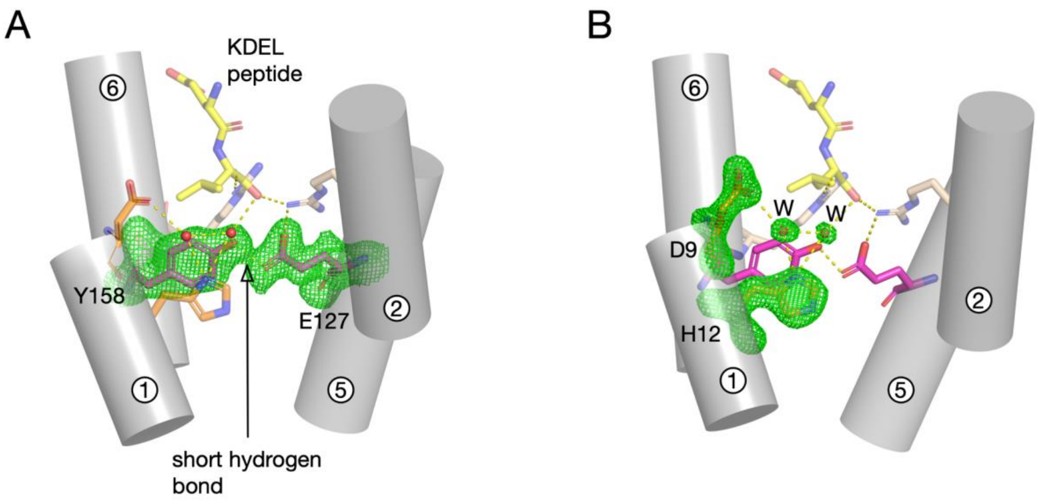
OMIT map showing the AEKDEL peptide captured in the receptor. A. Polder OMIT map (green mesh) contoured at 3σ, showing the position of the peptide in the KDELR, shown in cartoon (grey) in the plane of the membrane. **B.** Zoomed in view of the OMIT map shown in **A**. The density for the N-terminal threonine in the TAEKDEL ligand was too weak to model in the residue.

Fig. S8.



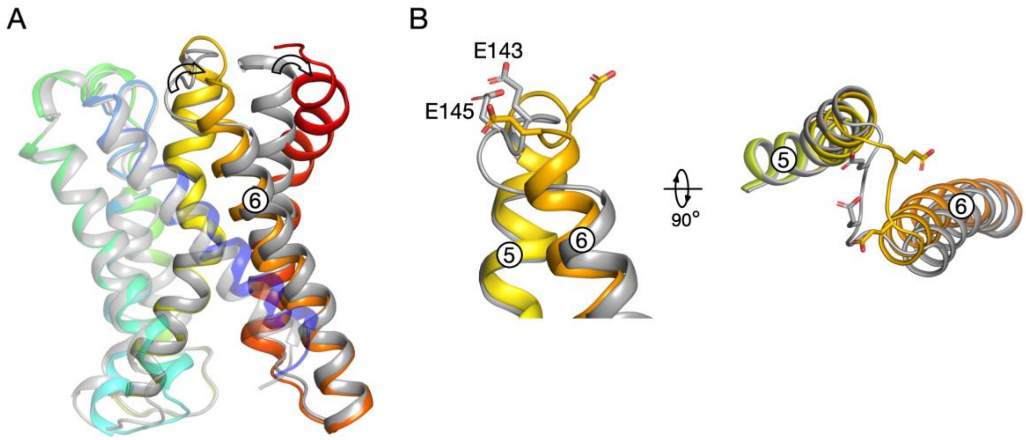
Comparison of receptor in peptide-bound and apo state. Plot showing the mean displacement of the center of each residue within the receptor upon peptide binding. The dashed boxes show regions of substantial movement and are highlighted on the structures below. The Apo structure is shown in Jones' rainbow (colored blue to red) representation while the peptide-bound structure is shown in grey. The TAEKDEL peptide is shown in yellow. The dashed red box around the TM6 label in the plot indicates the lesser extent of the helix in the Apo structure.

Fig. S9.



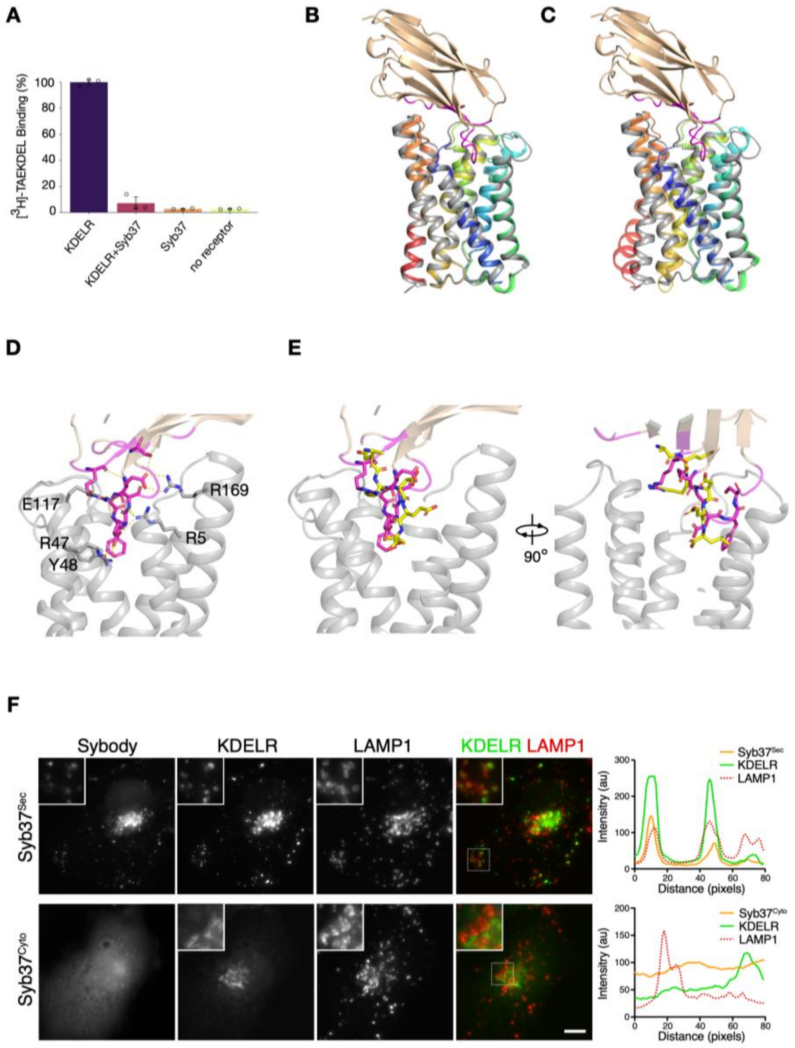
OMIT maps showing the short hydrogen bond interaction between Glu127 and Tyr158 upon peptide binding and water network. A. View of the short hydrogen bond formed between Glu127 and Tyr158 following binding of the TAEKDEL peptide (yellow sticks). The electron density from a Polder OMIT map is shown (green mesh), contoured at 3 σ, around Glu127 and Try158. Two water molecules are shown (red spheres), connecting Tyr158 to His12, suggesting a possible mechanism for pH sensing in the ER and Golgi. The second water molecule is observed forming a hydrogen bond to Asp9, mutation of which results in a receptor that is severely reduced in KDEL peptide binding (18). Hydrogen bonds shown as dashed lines (yellow). Arg47 and Arg159 (wheat sticks) are shown coordinating the carboxy terminus of the peptide. **B.** Equivalent view showing the Polder OMIT map (green mesh), contoured at 5 σ, around His12 and Asp9 and the two water molecules (W) linking these residues to Tyr158 and the carboxy terminus of the TAEKDEL peptide.

Fig. S10.



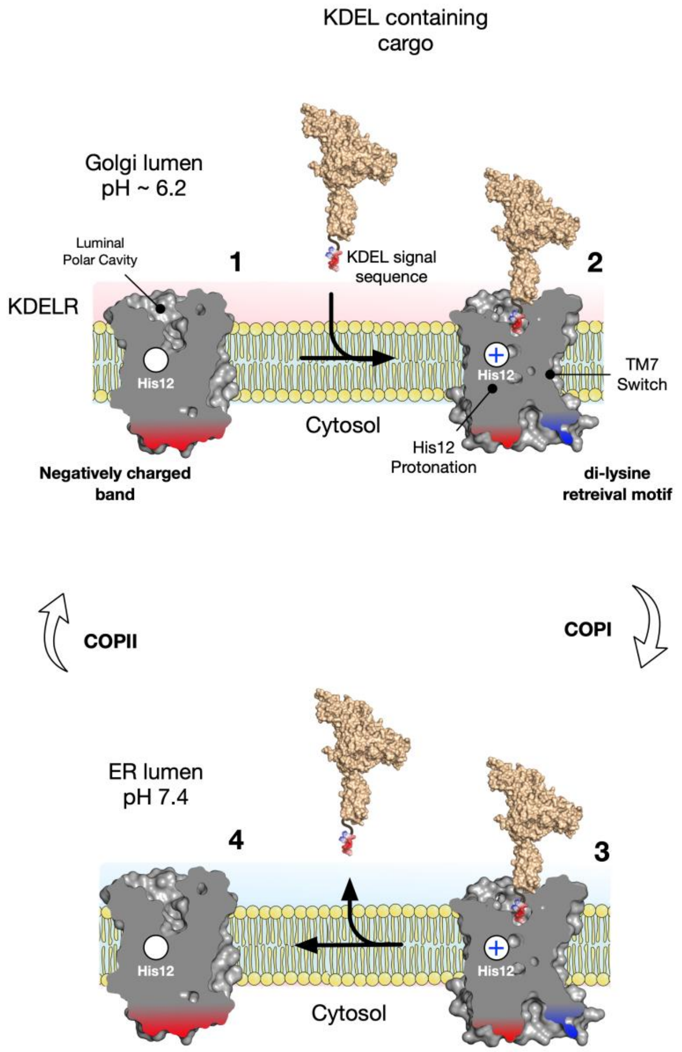
Structural comparison showing the movement of TM6 upon peptide binding. **A.** Overlay of the Apo (grey) and peptide-bound (colored blue to red) receptor structures with the movements in TM6 and TM7 indicated by arrows (open black). **B.** TM6 extends by one helical turn following the movement at the cytoplasmic end of TM7 upon peptide binding. This movement results in the repositioning of the conserved acidic side chains, Glu143 and Glu145, which contributes to the disruption of the negative electrostatic band projecting into the cytoplasm, and shown in Fig. S5.

Fig. S11.



Syb37 block KDEL peptide binding but does not induce exposure of the di-lysine retrieval motif. **A.** Bar chart showing binding of KDEL peptide to the receptor in detergent. **B.** Structural overlay of the KDELR (grey) in complex with Syb37 (wheat) with the apo KDELR structure (colored blue to red). The receptors overlay with an r.m.s.d. 0.73 Å. The variable regions of Syb37, including the CDR3 loop are shown in magenta. **C.** Equivalent overlay of the KDELR-Syb37 complex on the peptide bound receptor, which overlays with an r.m.s.d. 1.97 Å. **D**. Zoomed in view of the polar cavity showing the interactions made to Syb37. Hydrogen bonds are shown as dashed lines. **E.** Superposition of the TAEKDEL peptide structure onto the KDELR-Syb37 complex. The CDR3 loop (magenta sticks) occupies a similar position to the TAEKDEL peptide, but does not engage Arg159, instead interacting with Arg269 further along the helix. **F.** KDELR localization in the presence of ER and Golgi luminal $\text{Syb37}^{\text{Sec}}$ and cytoplasmic $\text{Syb37}^{\text{Cyto}}$. Cells were co-stained with the lysosomal marker LAMP1. Scale bar is 10µm. Enlarged inset regions and line graphs show the co-localization of KDELR and Syb37^{Sec} to LAMP1-positive structures.

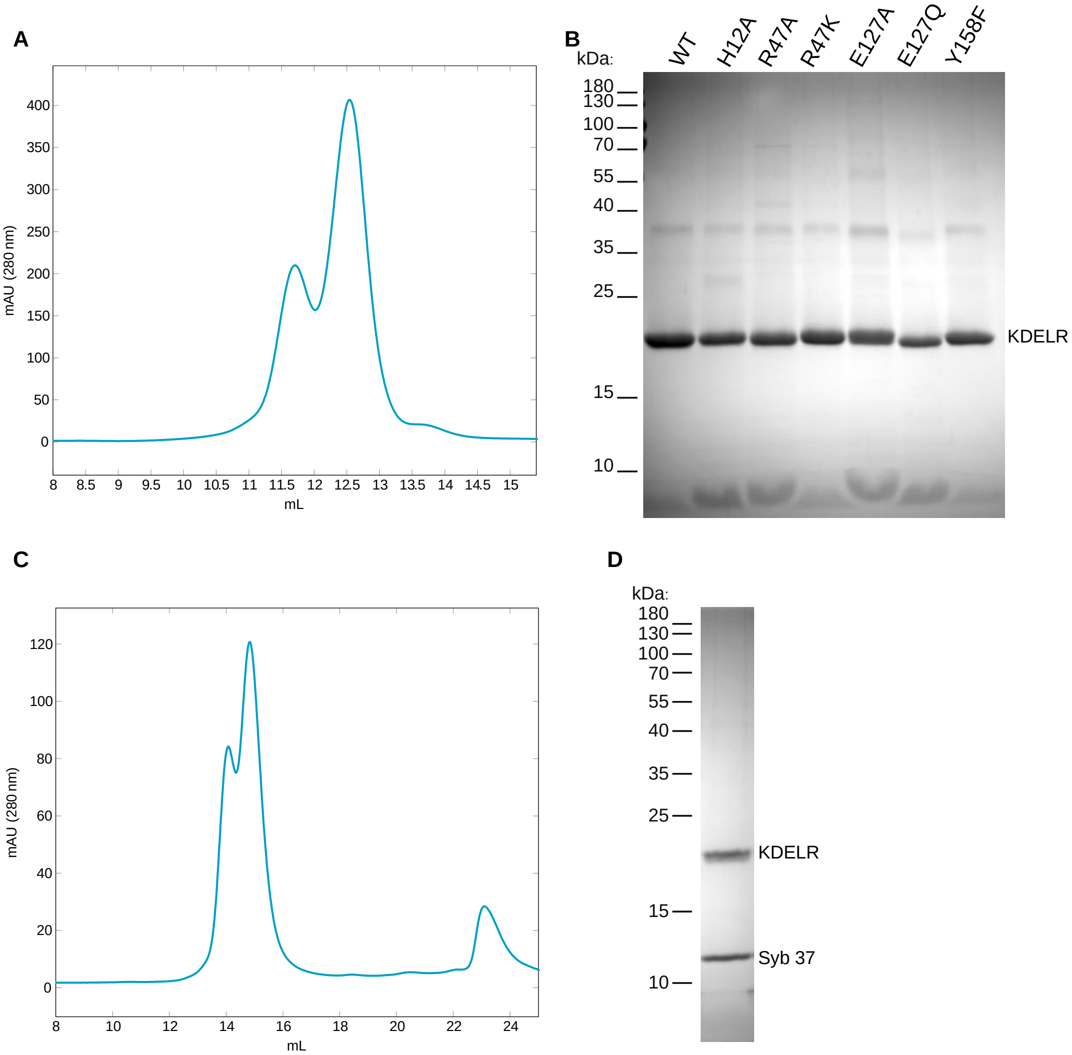
Fig. S12.



Model for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. Taken together, our data suggest a model for pH dependent retrieval of ER chaperones from the Golgi apparatus via the KDELR in mammalian cells. For Golgi to ER retrieval an empty receptor **(1)** must first bind to a KDEL protein in the luminal polar cavity at acidic pH **(2)**. Binding results in the movement of TM1 and TM6, with the position of TM6 being stabilized through the formation of a strong hydrogen bond interaction between Tyr158 on TM6 and Glu127 on TM5, which serves to lock the peptide in place. Formation of this lock is likely facilitated through the protonation of His12 on TM1, acting as the pH sensor for the system. The movement of TM6 causes an outward rotation of the C-terminal end of TM7, resulting in substantial rearrangement of the electrostatic surface of the receptor facing the cytosol. The cytoplasmic negatively charged band present in the apo receptor **(1)** gives way to a more basic motif on TM7, as the C-terminal lysine side chains are exposed, revealing the likely

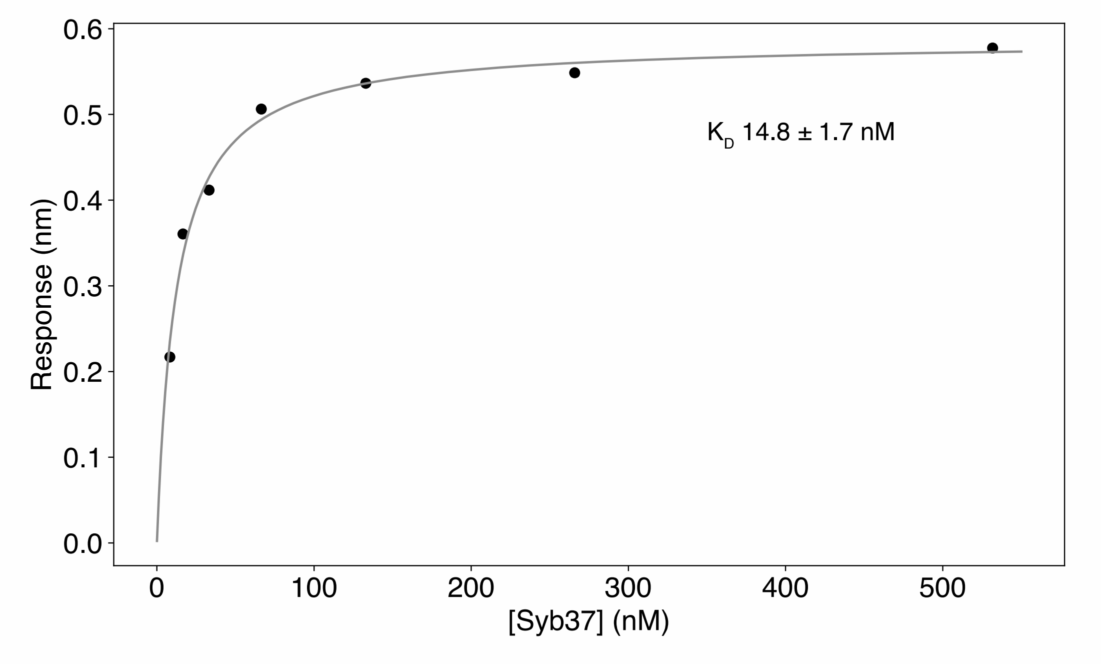
COPI binding site **(2)**. Following trafficking back to the ER, the change to neutral pH results in deprotonation of His12, causing TM1 and TM6 to move back and release the KDEL protein into the ER lumen **(3)**. As TM7 moves back, the C-terminal lysine motif packs against TM5, re-establishing the continuous negatively charged band, which may form part of a COPII binding motif **(4)**. The apo receptor is then trafficked back to the Golgi to repeat the cycle. The requirement for a pH modulated lock that requires the presence of the KDEL peptide to form, explains how this retrieval system is able to traffic ER chaperones against a millimolar concentration gradient using the extremely shallow pH gradient established between these two organelles. An important discovery from this work is that both the lysine COPI and putative acidic COPII recognition motifs are mutually exclusive, resulting in an elegant way for the KDEL receptor to switch between the ER to Golgi and Golgi to ER trafficking systems within the secretory pathway.

Fig. S13.



Purification of the KDELR and the KDELR-sybody complex. A. Size exclusion chromatography trace of the purified WT KDELR on an S200 (10/300) Superdex column. The two peaks likely correspond to monomer and dimer of the KDEL receptor. **B.** SDS-PAGE analysis of the pooled peaked for the WT and variants of the receptor**. C.** Size exclusion chromatography trace of the WT KDELR-Syb37 complex on an SRT-C SEC 300 SEC column. **D.** SDS-PAGE analysis of the SEC peak from **C.** showing formation of the complex.

Fig. S14.



Selection of a sybody with high affinity for the luminal face of the KDEL receptor. Representative curve for the binding of Syb37 to the *Gg*KDELR using bilayer interferometry. The K_D was calculated from the mean of three independent experiments with the standard deviation shown.

Table S1. Crystallographic data table. Data collection and refinement statistics.

	KDELR (PDB: 6I6B)	KDELR + TAEKDEL (PDB: 6I6H)	KDELR-Syb37 (PDB: 6I6J)
Data Collection			
Space Group	C 2 2 2_1	P 2_1	P 2_1 2_1 2_1
Cell Dimensions			
a, *b*, *c* (Å)	47.90, 103.75, 101.59	47.87, 37.50, 62.75	50.40, 52.98, 133.06
α, β, γ (°)	90, 90, 90	90, 95.42, 90	90, 90, 90
Wavelength (Å)	0.969	0.969	0.969
Resolution (Å)	51.88-2.59 (2.66-2.59)	47.66-2.00 (2.05-2.00)	47.13 - 2.23 (2.29-2.23)
R_{pim}	0.158 (0.348)	0.096 (0.895)	0.055 (0.536)
R_{merge}	0.359 (1.069)	0.282 (2.588)	0.122 (1.247)
$I/\sigma I$	4.1 (1.9)	6.8 (1.7)	7.0 (1.6)
Completeness (%)	99.9 (98.0)	99.2 (97.0)	99.5 (95.9)
Multiplicity	6.2 (6.3)	9.5 (9.1)	6.2 (6.3)
CC1/2	0.900 (0.519)	0.992 (0.721)	0.865 (0.536)
Wilson B-factor (Å^2)	23.7	22.6	40.9
Refinement			
No. reflections	50410	143906	111218
Unique Reflections	8183	15154	18108
R_{work}/R_{free}	0.24/0.27	0.19/0.23	0.21/0.25
No. atoms			
Protein	1680	1755	2638
(B_{res})*, protein atoms (Å^2)	30.8	25.5	49.1
(B_{res}), peptide atoms (Å^2)	n/a	36.4	n/a
(B_{res}), waters (Å^2)	27.7	44.3	48.4
(B_{res}), lipid atoms (Å^2)	51.0	65.5	67.9
Ramachandran favoured	97.0	97.6	97.5
Ramachandran outliers	0.0	0.0	0.0
R.M.S deviations			
Bond lengths (Å)	0.005	0.013	0.004
Bond angles (°)	0.96	1.49	0.95

* B_{res} are the residual B factors of the model.

Movie S1. Structural morph between the apo and peptide bound states of the KDEL receptor. The movie starts with the apo structure of KDEL receptor (coloured from blue to red) sitting vertically with respect to the membrane. The binding site is shown in sticks (coloured grey). The TAEKDEL peptide (coloured yellow), is then modeled entering the apo receptor. The receptor then morphs to show the crystal structure of the peptide bound state. Water molecules present in the peptide bound state are shown (red spheres). The TAEKDEL peptide is then modeled being released, and the structure relaxes back to the apo state. The view is then changed and zoomed in to observe the structural changes in the peptide binding site following the transition to the bound state. The interaction and length of the hydrogen bond between E127 and Y158 is shown. Residues in the binding site are labeled. The two waters coordinating His12 with Y158 and the carboxy terminus of the peptide are highlighted in green. Following the morph, a structural overlay is shown with the apo receptor overlaid on the peptide bound state (coloured grey), with binding site residues and peptide shown in sticks.

References and Notes

1. L. Ellgaard, A. Helenius, Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4**, 181–191 (2003). doi:10.1038/nrm1052 Medline

2. S. Munro, H. R. Pelham, A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48**, 899–907 (1987). doi:10.1016/0092-8674(87)90086-9 Medline

3. J. C. Semenza, K. G. Hardwick, N. Dean, H. R. Pelham, ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. *Cell* **61**, 1349–1357 (1990). doi:10.1016/0092-8674(90)90698-E Medline

4. F. M. Townsley, G. Frigerio, H. R. Pelham, Retrieval of HDEL proteins is required for growth of yeast cells. *J. Cell Biol.* **127**, 21–28 (1994). doi:10.1083/jcb.127.1.21 Medline

5. H. R. Pelham, K. G. Hardwick, M. J. Lewis, Sorting of soluble ER proteins in yeast. *EMBO J.* **7**, 1757–1762 (1988). doi:10.1002/j.1460-2075.1988.tb03005.x Medline

6. L. Feng, W. B. Frommer, Evolution of Transporters: The Relationship of SWEETs, PQ-loop, and PnuC Transporters. *Trends Biochem. Sci.* **41**, 118–119 (2016). doi:10.1016/j.tibs.2015.11.014 Medline

7. M. J. Lewis, H. R. Pelham, Ligand-induced redistribution of a human KDEL receptor from the Golgi complex to the endoplasmic reticulum. *Cell* **68**, 353–364 (1992). doi:10.1016/0092-8674(92)90476-S Medline

8. G. Griffiths, M. Ericsson, J. Krijnse-Locker, T. Nilsson, B. Goud, H. D. Söling, B. L. Tang, S. H. Wong, W. Hong, Localization of the Lys, Asp, Glu, Leu tetrapeptide receptor to the Golgi complex and the intermediate compartment in mammalian cells. *J. Cell Biol.* **127**, 1557–1574 (1994). doi:10.1083/jcb.127.6.1557 Medline

9. D. W. Wilson, M. J. Lewis, H. R. Pelham, pH-dependent binding of KDEL to its receptor in vitro. *J. Biol. Chem.* **268**, 7465–7468 (1993). Medline

10. M. M. Wu, J. Llopis, S. Adams, J. M. McCaffery, M. S. Kulomaa, T. E. Machen, H.-P. H. Moore, R. Y. Tsien, Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem. Biol.* **7**, 197–209 (2000). doi:10.1016/S1074-5521(00)00088-0 Medline

11. M. M. Wu, M. Grabe, S. Adams, R. Y. Tsien, H.-P. H. Moore, T. E. Machen, Mechanisms of pH regulation in the regulated secretory pathway. *J. Biol. Chem.* **276**, 33027–33035 (2001). doi:10.1074/jbc.M103917200 Medline

12. I. Majoul, M. Straub, S. W. Hell, R. Duden, H. D. Söling, KDEL-cargo regulates interactions between proteins involved in COPI vesicle traffic: Measurements in living cells using FRET. *Dev. Cell* **1**, 139–153 (2001). doi:10.1016/S1534-5807(01)00004-1 Medline

13. S. R. Pfeffer, Unsolved mysteries in membrane traffic. *Annu. Rev. Biochem.* **76**, 629–645 (2007). doi:10.1146/annurev.biochem.76.061705.130002 Medline

14. N. Gomez-Navarro, E. Miller, Protein sorting at the ER-Golgi interface. *J. Cell Biol.* **215**, 769–778 (2016). doi:10.1083/jcb.201610031 Medline

15. Y. Tao, L. S. Cheung, S. Li, J.-S. Eom, L.-Q. Chen, Y. Xu, K. Perry, W. B. Frommer, L. Feng, Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature* **527**, 259–263 (2015). doi:10.1038/nature15391 Medline

16. F. M. Townsley, D. W. Wilson, H. R. Pelham, Mutational analysis of the human KDEL receptor: Distinct structural requirements for Golgi retention, ligand binding and retrograde transport. *EMBO J.* **12**, 2821–2829 (1993). doi:10.1002/j.1460-2075.1993.tb05943.x Medline

17. C. Barlowe, Signals for COPII-dependent export from the ER: What's the ticket out? *Trends Cell Biol.* **13**, 295–300 (2003). doi:10.1016/S0962-8924(03)00082-5 Medline

18. H. J. Sharpe, T. J. Stevens, S. Munro, A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158–169 (2010). doi:10.1016/j.cell.2010.05.037 Medline

19. J. L. Parker, S. Newstead, Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature* **551**, 521–524 (2017). Medline

20. A. A. Scheel, H. R. Pelham, Identification of amino acids in the binding pocket of the human KDEL receptor. *J. Biol. Chem.* **273**, 2467–2472 (1998). doi:10.1074/jbc.273.4.2467 Medline

21. F. Letourneur, E. C. Gaynor, S. Hennecke, C. Démollière, R. Duden, S. D. Emr, H. Riezman, P. Cosson, Coatomer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. *Cell* **79**, 1199–1207 (1994). doi:10.1016/0092-8674(94)90011-6 Medline

22. L. P. Jackson, M. Lewis, H. M. Kent, M. A. Edeling, P. R. Evans, R. Duden, D. J. Owen, Molecular basis for recognition of dilysine trafficking motifs by COPI. *Dev. Cell* **23**, 1255–1262 (2012). doi:10.1016/j.devcel.2012.10.017 Medline

23. I. Zimmermann, P. Egloff, C. A. J. Hutter, F. M. Arnold, P. Stohler, N. Bocquet, M. N. Hug, S. Huber, M. Siegrist, L. Hetemann, J. Gera, S. Gmür, P. Spies, D. Gygax, E. R. Geertsma, R. J. P. Dawson, M. A. Seeger, Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife* **7**, e34317 (2018). doi:10.7554/eLife.34317 Medline

24. J. L. Parker, S. Newstead, Method to increase the yield of eukaryotic membrane protein expression in *Saccharomyces cerevisiae* for structural and functional studies. *Protein Sci.* **23**, 1309–1314 (2014). doi:10.1002/pro.2507 Medline

25. D. Drew, S. Newstead, Y. Sonoda, H. Kim, G. von Heijne, S. Iwata, GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat. Protoc.* **3**, 784–798 (2008). doi:10.1038/nprot.2008.44 Medline

26. M. Caffrey, V. Cherezov, Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731 (2009). doi:10.1038/nprot.2009.31 Medline

27. G. Winter, xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2009). doi:10.1107/S0021889809045701

28. G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter, G. Evans, DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. D* **74**, 85–97 (2018). doi:10.1107/S2059798317017235 Medline

29. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013). doi:10.1107/S0907444913000061 Medline

30. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi:10.1107/S0021889807021206 Medline

31. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991). doi:10.1107/S0108767390010224 Medline

32. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010). doi:10.1107/S0907444910007493 Medline

33. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010). doi:10.1107/S0907444909052925 Medline

34. E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea, G. Bricogne, Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004). doi:10.1107/S0907444904016427 Medline

35. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi:10.1038/nmeth.2019 Medline

Mit der Dummheit kämpfen Götter selbst vergebens.
— Friedrich Schiller